

 Open access • Journal Article • DOI:10.1016/J.JEBO.2016.10.020

Heterogeneous peer effects in education — [Source link](#)

[Eleonora Patacchini](#), [Edoardo Rainone](#), [Yves Zenou](#)

Institutions: [Monash University](#)

Published on: 01 Feb 2017 - [Journal of Economic Behavior and Organization](#) (North-Holland)

Topics: [Educational attainment](#) and [Convergence \(relationship\)](#)

Related papers:

- [Identification of Endogenous Social Effects: The Reflection Problem](#)
- [Identification of peer effects through social networks.](#)
- [Peer Effects and Social Networks in Education](#)
- [Heterogeneous Peer Effects in Education](#)
- [A Social Interactions Model with Endogenous Friendship Formation and Selectivity](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/heterogeneous-peer-effects-in-education-o8x8ttbtw>



BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Heterogeneous peer effects in education

by Eleonora Patacchini, Edoardo Rainone and Yves Zenou

January 2016

Number

1048



BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Heterogeneous peer effects in education

by Eleonora Patacchini, Edoardo Rainone and Yves Zenou

Number 1048 - January 2016

The purpose of the Temi di discussione series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

Editorial Board: PIETRO TOMMASINO, PIERGIORGIO ALESSANDRI, VALENTINA APRIGLIANO, NICOLA BRANZOLI, INES BUONO, LORENZO BURLON, FRANCESCO CAPRIOLI, MARCO CASIRAGHI, GIUSEPPE ILARDI, FRANCESCO MANARESI, ELISABETTA OLIVIERI, LUCIA PAOLA MARIA RIZZICA, LAURA SIGALOTTI, MASSIMILIANO STACCHINI.

Editorial Assistants: ROBERTO MARANO, NICOLETTA OLIVANTI.

ISSN 1594-7939 (print)

ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

HETEROGENEOUS PEER EFFECTS IN EDUCATION

by Eleonora Patacchini[§], Edoardo Rainone* and Yves Zenou[#]

Abstract

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

JEL Classification: C31, D85, Z13.

Keywords: spatial autoregressive model, heterogeneous spillovers, 2SLS estimation, Bayesian estimation, education.

Contents

| | |
|--------------------------------------------------|----|
| 1. Introduction..... | 5 |
| 2. Data description | 8 |
| 3. Empirical model and estimation strategy | 10 |
| 4. Estimation results..... | 13 |
| 5. Understanding the mechanisms | 14 |
| 6. Robustness checks | 21 |
| 7. Short-run versus long-run effects | 27 |
| 8. Concluding remarks..... | 29 |
| References | 30 |
| Appendices | 36 |
| Figures and tables..... | 53 |

[§] Cornell University, EIEF and CEPR. E-mail: ep454@cornell.edu

*Bank of Italy. E-mail: edoardo.rainone@bancaditalia.it.

[#] Stockholm University, Research Institute of Industrial Economics (IFN) and GAINS. E-mail: yves.zenou@ne.su.se

1 Introduction

The influence of peers on education outcomes has been widely studied both in economics and sociology (Sacerdote, 2011). Yet many questions remain unanswered.¹ In particular, very little is known about the effect of school peers on the long-run outcomes of students. This is primarily due to the absence of information on peers during teenage years together with long-run outcomes of individuals in most existing data. Besides, the mechanisms by which the peer effects affect education are unclear.

In this paper, we provide an analysis on the long-run effects of high-school peers on years of schooling and put forward the role of different types of ties for educational outcomes.

For that, we use the unique information on friendship networks² among students in the United States provided by the Addhealth data. We exploit three unique features of the AddHealth data: (i) the nomination-based friendship information, which allows us to reconstruct the precise geometry of social contacts during high-school years, (ii) the variation in friendship network topology between Wave I and Wave II, which enables us to distinguish between *short-lived ties* and *long-lived ties* and (iii) the longitudinal dimension, which provides information about each individual and her/his friends' outcomes during the adulthood.

More specifically, we use the different waves of the AddHealth data by looking at the impact of school friends nominated in the first two waves in 1994-1995 and in 1995-1996 on own educational outcome (when adult) reported in the fourth wave in 2007-2008 (measured by the number of completed years of full time education). We define a *long-lived tie* relationship between two students if they have nominated each other in both waves (i.e. in Wave I in 1994-1995 and in Wave II in 1995-1996) and a *short-lived tie* relationship if they have nominated each other in one wave only.

The existing (limited) studies on heterogeneous peer effects have looked at peer effects for different types of individuals (e.g. males versus females, whites versus blacks).³ In doing so, they assume homogeneity within each type. Our goal is to study the individual response to heterogeneous types of bilateral relationships by considering the case of *heterogeneous links*. Previous work avoided this situation because of technical difficulties in the estimation

¹The constraints imposed by the available disaggregated data force many studies to analyze peer effects in education at a quite aggregate and arbitrary level, such as at the high school (Evans et al., 1992), the census tract (Brooks-Gunn et al., 1993), and the ZIP code level (Datcher, 1982; Corcoran et al., 1992) where individuals reside. The importance of peer effects as distinct from neighborhood influences is still a matter of debate in many fields (see, e.g., the literature surveys by Durlauf, 2004, Ioannides and Topa, 2010, and Ioannides, 2011, 2012).

²The economics of networks is a growing field. For overviews, see Jackson (2008), Blume et al. (2011), Ioannides (2012), Boucher and Fortin (2015), Graham (2015), and Jackson and Zenou (2015).

³See e.g. Griffith and Rask (2014), Tincani (2015) and Yakusheva et al. (2014).

of heterogeneous endogenous peer effects.

In this paper, we extend the Liu and Lee (2010) 2SLS approach to a network model with different interaction matrices. The asymptotic consistency and efficiency of the proposed estimators are proved. We also employ a Bayesian inferential methods to integrate a network formation model with the study of behavior over the formed networks. Finally, we consider possible measurement errors in peer groups using a simulation experiment. Our results are robust to various types of network topology misspecification. This study is the first to estimate heterogeneous spillovers in a network model. We do so using both a frequentist and a bayesian approach.

Our results show that there are strong and persistent peer effects in education. When looking at the role of *short-lived* and *long-lived ties* in educational decisions, it appears that the education decisions of *short-lived* ties have no significant effect on individual long-run outcomes, regardless of whether peers are interacting in lower or higher grades. On the contrary, we find that the educational choices of *long-lived ties* have a positive and significant effect on own educational outcome.

There is a large literature on the role of different ties in the labor market. In particular, Granovetter (1973, 1974, 1983) initiated a strand of studies looking at the effects of *weak* versus *strong* ties. Strong ties are viewed as *stable* relationships and weak ties as *unstable* relationships.⁴ Interestingly, compared to the literature on the labor market, we find the opposite result for education outcomes.^{5,6} Indeed, we show that stable rather than unstable ties matter for education. This is reasonable given that outcomes and mechanisms are different in the two contexts. While random encounters may be helpful in providing information about jobs, they typically do not contribute to shape social norms, values and attitudes (see, e.g. Coleman, 1988, Wellman and Wortley, 1990). The collective value of “social networks”, which is a relevant driver of long-run influences, need time and repeated interactions to be established (Putnam, 2000).

⁴In his seminal papers, Granovetter defines *weak ties* in terms of lack of overlap in personal networks between any two agents, i.e. weak ties refer to a network of acquaintances who are less likely to be socially involved with one another. Formally, two agents A and B have a weak tie if there is little or no overlap between their respective personal networks. Vice versa, the tie is *strong* if most of A’s contacts also appear in B’s network.

⁵Yakubovich (2005) uses a large scale survey of hires made in 1998 in a major Russian metropolitan area and finds that a worker is more likely to find a job through weak ties than through strong ties. These results come from a within-agent fixed effect analysis, so they are independent of workers’ individual characteristics. Using data from a survey of male workers from the Albany NY area in 1975, Lin et al. (1981) find similar results. Lai et al. (1998) and Marsden and Hurlbert (1988) also find that weak ties facilitate reaching a contact person with higher occupational status who, in turn, leads to better jobs, on average.

⁶See also Patacchini and Zenou (2008) who find evidence of the strength of weak ties in crime.

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

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

There are very few studies looking at the long-run effects of friendship on human capital accumulation. Using the Wisconsin Longitudinal Study of Social and Psychological Factors in Aspiration and Attainment (WLS), Zax and Rees (2002) were the first to analyze the role of friendships in school on future earnings. Using the AddHealth data, Bifulco et al. (2011) study the effect of school composition (percentage of minorities and college educated mothers among the students in one's school cohort) on high-school graduation and post-secondary outcomes.

The paper unfolds as follows. Our data are described in Section 2, while the estimation

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

2 Data description

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth). The AddHealth survey has been designed to study the impact of the social environment (i.e. friends, family, neighborhood and school) on adolescents' behavior in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in the years 1994-1995 (Wave I). Every pupil attending the sampled schools on the interview day is asked to compile a questionnaire (*in-school data*) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. A subset of adolescents selected from the rosters of the sampled schools, about 20,000 individuals, is then asked to compile a longer questionnaire containing more sensitive individual and household information (*in-home and parental data*). Those subjects are interviewed again in 1995-1996 (Wave II), in 2001-2002 (Wave III), and in 2007-2008 (Wave IV).

From a network perspective, the most interesting aspect of the AddHealth data is the friendship information, which is based upon actual friends nominations. Indeed, pupils were asked to identify their best friends from a school roster (up to five males and five females).⁷ This information is collected in Wave I and one year after, in Wave II. As a result, one can reconstruct the whole geometric structure of the friendship networks and their evolution, at least in the short run. Such detailed information on social interaction patterns allows us to measure the peer group more precisely than in previous studies by knowing exactly who nominates whom in a network (i.e. who interacts with whom in a social group).

Moreover, and this has not been done before, one can distinguish between *long-lived* and

⁷The limit in the number of nominations is not binding (even by gender). Less than 1% of the students in our sample show a list of ten best friends, both in Wave I and Wave II.

short-lived ties in the data. We define a *long-lived* friendship between two students if they have nominated each other in both waves (i.e. in Wave I in 1994-1995 *and* in Wave II in 1995-1996) and a *short-lived* friendship if they have nominated each other in one wave only (Wave I *or* Wave II).

By matching the identification numbers of the friendship nominations to respondents' identification numbers, one can also obtain information on the characteristics of nominated friends. In addition, the longitudinal structure of the survey provides information on both respondents and friends during adulthood. In particular, the questionnaire of Wave IV contains detailed information on the highest education qualification achieved. We measure *educational attainment* in completed years of full time education in Wave IV.⁸ Social contacts (i.e. friendship nominations) are, instead, collected in Waves I and II.

Our final sample of in-home Wave I students (and friends) that are followed over time and have non-missing information on our target variables both in Waves I, II and IV consists of 1,819 individuals distributed over 116 networks. This large reduction in sample size with respect to the original sample is mainly due to the network construction procedure - roughly 20 percent of the students do not nominate any friends and another 20 percent cannot be correctly linked. In addition, we exclude networks consisting of 2-3 individuals, those with more than 400 members and individuals who are not followed in Wave IV.⁹ In Wave I, the mean and the standard deviation of network size are roughly 9.5 and 15, respectively. Roughly 61% of the nominations are not renewed in Wave II, and about 44% new ones are made. On average, these adolescents have roughly 30% long-lived ties and 70% short-lived ties. Further details on nomination data can be found in Table A1 in Appendix A. Appendix A also gives a precise definition of the variables used in our study as well as their descriptive statistics (see Table A1).¹⁰

⁸More precisely, the Wave IV questionnaire asks about the highest education qualification achieved (distinguishing between 8th grade or less, high school, vocational/technical training, bachelor's degree, graduate school, master's degree, graduate training beyond a master's degree, doctoral degree, post baccalaureate professional education). Those with high school qualifications and higher are also asked to report the exact year when the highest qualification was achieved. Such information allows us to construct a reliable measure of each individual's completed years of education.

⁹We do not consider networks at the extremes of the network size distribution (i.e. consisting of 2-3 individuals or more than 400) because peer effects can show extreme values (too high or too low) in these edge networks (see Calvó-Armengol et al., 2009). The representativeness of the sample is preserved. Summary statistics are available upon request.

¹⁰Information at the school level, such as school quality and the teacher/pupil ratio, is also available but we do not need to use it since our sample of networks is within schools and we use fixed network effects in our estimation strategy.

3 Empirical model and estimation strategy

3.1 Empirical model

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

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

where $y_{i,r,t+1}$ is the highest education level reached by individual i at time $t+1$ who belonged to network r at time t , where time $t+1$ refers to Wave IV in 2007-2008 while time t refers to Wave I in 1994-1995 and/or Wave II in 1995-1996 (depending on whether we consider short-lived or long-lived ties). Similarly, $y_{j,r,t+1}$ is the highest education level reached by individual j at time $t+1$ who has been nominated as his/her friend by individual i at time t in network r . Furthermore, $\mathbf{x}'_{i,r,t,t+1} = (x_{i,r,t,t+1}^1, \dots, x_{i,r,t,t+1}^m)'$ indicates the M variables accounting for observable differences in individual characteristics of individual i both at times t (e.g. self esteem, mathematics score, quality of the neighborhood, etc.) and $t+1$ (marital status, age, children, etc.) of individual i . Some characteristics are clearly the same at times t and $t+1$, such as race, parents' education, gender, etc. Also $g_{i,r,t}^L = \sum_{j=1}^{n_r} g_{ij,r,t}^L$ and $g_{i,r,t}^S = \sum_{j=1}^{n_r} g_{ij,r,t}^S$ are the total number of long-lived and short-lived friends each individual i has in network r at time t . Finally, $\epsilon_{i,r}$'s are i.i.d. innovations with zero mean and variance σ^2 for all i and r .

Let $\mathbf{Y}_r = (y_{1,r,t+1}, \dots, y_{n_r,r,t+1})'$, $\mathbf{X}_r = (x_{1,r,t,t+1}, \dots, x_{n_r,r,t,t+1})'$, and $\epsilon_r = (\epsilon_{1,r}, \dots, \epsilon_{n_r,r})'$. Denote the $n_r \times n_r$ adjacency matrix by $\mathbf{G}_r = [g_{ij,r}]$, the row-normalized of \mathbf{G}_r by \mathbf{G}_r^* , and the n_r -dimensional vector of ones by $\mathbf{1}_{n_r}$. As above, let us split the adjacency matrix into two submatrices \mathbf{G}_r^L and \mathbf{G}_r^S , which keep trace of long-lived and short-lived friends, respectively.

Then, model (1) can be written in matrix form as:

$$\mathbf{Y}_r = \phi^L \mathbf{G}_r^L \mathbf{Y}_r + \phi^S \mathbf{G}_r^S \mathbf{Y}_r + \mathbf{X}_r^* \beta + \eta_r \mathbf{1}_{n_r} + \epsilon_r, \quad (2)$$

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

For a sample with \bar{r} networks, stack up the data by defining $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_{\bar{r}})'$, $\mathbf{X}^* = (\mathbf{X}'_1, \dots, \mathbf{X}'_{\bar{r}})'$, $\epsilon = (\epsilon'_1, \dots, \epsilon'_{\bar{r}})'$, $\mathbf{G} = \text{D}(\mathbf{G}_1, \dots, \mathbf{G}_{\bar{r}})$, $\mathbf{G}^* = \text{D}(\mathbf{G}_1^*, \dots, \mathbf{G}_{\bar{r}}^*)$, $\iota = \text{D}(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_{\bar{r}}})$ and $\eta = (\eta_1, \dots, \eta_{\bar{r}})'$, where $\text{D}(\mathbf{A}_1, \dots, \mathbf{A}_K)$ is a block diagonal matrix in which the diagonal blocks are $n_k \times n_k$ matrices \mathbf{A}_k 's. For the entire sample, the model is thus:

$$\mathbf{Y} = \phi^L \mathbf{G}^L \mathbf{Y} + \phi^S \mathbf{G}^S \mathbf{Y} + \mathbf{X}^* \beta + \iota \cdot \eta + \epsilon. \quad (3)$$

In this model, ϕ^L and ϕ^S represent *the endogenous effects*, i.e. the agent's outcome depends on that of his/her friends, while γ^L and γ^S represent *the contextual effect*, i.e. the agent's choice/outcome depends on the exogenous characteristics of his/her friends. The vector of network fixed effects η captures *the correlated effect* where agents in the same network may behave similarly as they have similar unobserved individual characteristics or they face a similar (e.g. institutional) environment.¹¹

3.2 Identification and estimation

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

Reflection problem In linear-in-means models, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of peers' choice of effort (*endogenous effects*) and peers' characteristics (*contextual effects*) that do have an impact on their effort choice (the so-called *reflection problem*; Manski, 1993). Basically, the reflection problem arises because, in the standard approach, individuals interact in groups, that is individuals are affected by all individuals belonging to their group and by nobody outside the group. In the case of social networks, instead, this is nearly never true since the reference group is individual specific. For example, take individuals i and k such

¹¹As an analogy with time series models, the model in (3) can be referred to as a SARARMA(p, q) with $p = 0$ and $q = 2$, where p and q are the maximum number of spatial lags for the error and the outcome, respectively.

that $g_{ik} = 1$. Then, individual i is directly influenced by $g_i = \sum_{j=1}^{n_i} g_{ij}y_j$ while individual k is directly influenced by $g_k = \sum_{j=1}^{n_k} g_{kj}y_j$, and there is little chance for these two values to be the same unless the network is complete (i.e. everybody is linked with everybody).¹²

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

Sorting In most cases, individuals sort into groups non-randomly. For example, kids whose parents are low educated or worse than average in unmeasured ways would be more likely to sort with low human capital peers. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are major sources of bias. The richness of social network data (where we observe individuals over networks) provides a possible way out by the use of *network fixed effects*. Network fixed effects are a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 2).¹³

To summarize, in this paper, we consider 2SLS estimators (Liu and Lee, 2010) with network fixed effects and propose two innovations. First, we use two centralities, one for

¹²Formally, social effects are identified (i.e. no reflection problem) if $\mathbf{G}^2 \neq \mathbf{0}$, where \mathbf{G}^2 keeps track of indirect connections of length 2 in the network. This means that we need at least two individuals in the networks that have different links. This condition is generally satisfied in every real-world network.

¹³93% of networks have a size below 35. As a robustness check, in Section 6.1 below, we will also consider an explicit model of network formation and estimate simulatenously the outcome equation (1) and the network formation equation.

long-lived ties and one for short-lived ties. Second, we take advantage of the longitudinal structure of our data and only include values lagged in time in the different instrumental matrices (i.e. observed in Wave I). Appendix B reviews the approach proposed by Liu and Lee (2010) and highlights the modification that is implemented in this paper.¹⁴

4 Estimation results

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

4.1 Long-run peer effects

Table 1 collects the estimation results of model (1), without distinguishing between long-lived and short-lived ties. In other words, students i and j are friends, i.e. $g_{ij} = 1$, if they have nominated each other in Wave I. We then look at the impact of friends from Wave I on own educational attainment in Wave IV. The first three columns show the results when using the traditional set of instruments whereas, in the last three columns, the instrumental set only contains variables lagged in time (see Appendix B). The first-stage partial F-statistics (Stock et al., 2002 and Stock and Yogo, 2005) reveals that our instruments are quite informative and the OIR test provides evidence in line with their validity. Table 2 shows the results of the first step. The results in Table 1 do not change to any considerable extent across columns and reveal that the effect of friends' education on own education is always significant and positive, i.e., there are *long-lived and persistent peer effects in education*. This shows that the "quality" of friends (in terms of future educational achievement) from high school has a positive and significant impact on the own future educational level, even though it might be that individuals who were close friends in 1994-1995 (Wave I) might no longer be friends in 2007-2008 (Wave IV). According to the bias-corrected 2SLS estimator,¹⁵ in a group of two friends, a standard deviation increase in the years of education of the friend translates into

¹⁴Observe that most of the traditional problems in the identification of peer effects arises when outcome and linking decisions are taken at the same time. In our analysis, we do not have this problem since there is a time lag between when friends are chosen (Waves I and II in 1994-1996) and when the outcome (education) is observed (Wave IV in 2007-2008). In addition, the longitudinal aspect of our analysis provides IVs at different points in time, i.e. the characteristics of indirect peers when they are at school and when they are adults. It is thus possible to use only variables lagged in time as instruments to ensure that the instruments are not correlated with the contemporaneous error term.

¹⁵The bias-corrected 2SLS estimator is our preferred one since we have relatively small networks (see Appendix C).

a roughly 5.4 percent increase of a standard deviation in the individual years of education (roughly two more months of education). If we consider an average group of four best friends (linked to each other in a network), a standard deviation increase in the level of education of each of the peers translates into a roughly 16 percent increase of a standard deviation in the individual’s educational attainment (roughly seven more months of education). This is a non-negligible effect, especially given our long list of controls and the fact that friendship networks might have changed over time. The influence of peers at school seems to be carried over time.

4.2 The role of long-lived ties

We would now like to determine how long- and short-lived ties affect educational choices by estimating the magnitude of ϕ^L and ϕ^S in equation (1). Table 3 shows the estimation results of model (1).¹⁶ We find that the educational choices of short-lived friends have *no* significant impact on individual educational outcomes (years of schooling) while the educational choices of long-lived friends do have a positive and significant effect on own educational outcome. In terms of magnitude, a standard deviation increase in aggregate years of education of peers nominated both in Waves I and II (long-lived friends) translates into roughly a 21 percent increase of a standard deviation in the individual’s educational attainment (roughly 8.3 more months of education). In an average group of four best friends (linked to each other in a network), a standard deviation increase of each of the peers translates into two more years of education. This is quite an important effect. It suggests that *long-lived friends* rather than *short-lived friends* matter for educational outcomes in the long run.¹⁷

5 Understanding the mechanisms

Our empirical results displayed in Table 3 suggest that the distinction between long-lived and short-lived friends is important for understanding long-run peer effects in education. In Section 5.1, we propose a simple theoretical model explaining this evidence. The idea underlying the theoretical mechanism is that convergence of preferences and formation of social norms need long-term relationships between peers. In Section 5.2, we provide additional empirical evidence ruling out alternative explanations.

¹⁶We show the results for the bias-corrected 2SLS estimator, with the traditional set of instruments and when the instrumental set only contains variables lagged in time. The qualitative results when using the alternative estimators in Appendix C remain qualitatively unchanged. The latter are available upon request.

¹⁷When estimating model (1) including only long-lived ties (i.e. $\mathbf{G}^S = 0$), we obtain comparable results.

5.1 Theoretical framework

In order to understand how long-lived and short-lived ties influence long-run educational outcomes, we extend the DeGroot (1974) model as follows.¹⁸ Consider a society consisting of a finite set of individuals $N = \{1, 2, \dots, n\}$ who are linked in a directed network and who would like to gather information about an unknown parameter θ . In our context, assume that there are two states of the world so that θ can be equal to either: {It is worth continuing studying} or {It is not worth continuing studying}. What is key in the DeGroot model is that agents update their beliefs by repeatedly taking *weighted averages* of their neighbors' beliefs (where neighbors are the people directly linked to each individual) with p_{ij} being the weight that agent i places on the current belief of agent j in forming his or her belief for the next period. If the network is strongly connected and at least some individuals listen to themselves, then, in the limit, everybody belonging to the same network will converge to a consensus and the influence of each person will depend on their position in the network. This means that, if there are *repeated interactions* between students from the same network, then they will all adopt the same social norm, which could be either {It is worth continuing studying} or {It is not worth continuing studying} depending on what the "influential" students believe about the social norm.¹⁹

We use this theoretical framework to understand why, in our empirical results, short-lived friends have *no* impact on own education decision while long-lived friends have. We define *short-lived* friends as students who interact with each other only *once* while *long-lived* friends where students who interact *for a longer time*. There are two types of friend relationships between students in a network \mathbf{G} : *short-lived* friendships ($l = S$) and *long-lived* friendships ($l = L$). Quite naturally, we assume that each agent has a long-lived relationship with him/herself, i.e. $g_{ii}^L = 1$ and $g_{ii}^S = 0$. As in Section 3, this implies two types of adjacency matrices: $\mathbf{G}^L = \{g_{ij}^L\}$, where $g_{ij}^L = 1$ if i and j are long-lived friends and $\mathbf{G}^S = \{g_{ij}^S\}$, where $g_{ij}^S = 1$ if i and j are short-lived friends, with $\mathbf{G}^L + \mathbf{G}^S = \mathbf{G}$. Denote by $\tilde{\mathbf{G}}^L$ and $\tilde{\mathbf{G}}^S$ the row-normalized matrices of \mathbf{G}^L and \mathbf{G}^S , respectively. Because *short-lived* friendships only interact once, they will influence the beliefs of each other only in the initial period. The updating stops there since *short-lived* friends do not meet anymore and thus students only update their beliefs once. On the contrary, *long-lived* friends interact repeatedly and thus

¹⁸Appendix D contains the technical details of the DeGroot model.

¹⁹Because students have parents with different incomes or students have different costs of studying (some like to study while others don't), we can explain why, within a network with the same social norm, students take different decisions concerning the number of years they will spend in college. In particular, the students whose parents have low income or the students who have a high disutility of studying will end up not going to college even if the social norm in their network of friends says that it is worth continuing studying.

update their beliefs all the time as in the standard DeGroot model. Then all students in the network will therefore reach a consensus after many interactions, even though only long-lived friends will matter in the long run.

How do we solve this model? In the first period, both short-lived and long-lived students influence each other so that

$$\mathbf{b}^{(1)} = \tilde{\mathbf{G}} \mathbf{b}^{(0)}$$

where $\mathbf{b}^{(t)}$ is the vector of beliefs of *all students* at time t (i.e. both short-lived and long-lived students) and $\tilde{\mathbf{G}}$ is the row-normalized matrix of \mathbf{G} . Now relabel $\mathbf{b}^{(1)}$ as $\tilde{\mathbf{b}}^{(0)}$, i.e. $\tilde{\mathbf{b}}^{(0)} := \mathbf{b}^{(1)}$. We are now in the framework of the DeGroot model where the initial beliefs are given by $\tilde{\mathbf{b}}^{(0)}$. As a result, we can apply Proposition 8 in Appendix D. If the network $\tilde{\mathbf{G}}^L$ is strongly connected and if at least some agents pay attention to themselves, i.e. $g_{ii} > 0$ for some i , then all students will reach a consensus in the long run, which is determined by:

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

In equation (4), we see clearly the distinct influence of short-lived and long-lived friends. The updating matrix $\tilde{\mathbf{G}}^L$ is only depending on long-lived friends because they are the ones who interact over time and help reach a consensus. However, the initial beliefs are a function of the beliefs of both short-lived and long-lived friends since $\tilde{\mathbf{G}}$ includes both $\tilde{\mathbf{G}}^L$ and $\tilde{\mathbf{G}}^S$. In equation (4), we assume that all students have both long-lived and short-lived friends²⁰ so that they are all included in the convergence process $(\mathbf{b})^\infty$.

To illustrate this result, consider the example given at the end of Appendix D where the network is displayed in Figure A1. The adjacency matrix \mathbf{G} and the row-normalized one $\tilde{\mathbf{G}}$ are given by:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{G}} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Assume that both agents 1 and 2 and agents 1 and 3 have a *long-lived* friendship while agents 2 and 3 have a *short-lived* friendship. As stated above, we assume that each

²⁰Observe that if some students have only short-lived friends, then there will be no consensus. Indeed, even if just one student i has only short-lived friends, then $g_{ii}^L = 1$ and $g_{ij}^L = 0$ for all other j . Hence network $\tilde{\mathbf{G}}^L$ will not be strongly connected, and the baseline deGroot model will not work properly (i.e. depending on the network structure there will either be more than one component with a different “consensus” each, or no consensus at all).

agent has a long-lived relationship with him/herself, i.e. $g_{ii}^L = 1$ and $g_{ii}^S = 0$. We have:

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

Let us row-normalize these matrices so that

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

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

Let us first determine the initial beliefs. We have:

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

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

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

so that there is convergence to the following consensus:

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

In other words, no matter what beliefs $\tilde{\mathbf{b}}^{(0)}$ the agents start with, they all end up with

limiting beliefs corresponding to the entries of $(\mathbf{b})^\infty = \lim_{t \rightarrow \infty} (\tilde{\mathbf{G}}^L)^t \tilde{\mathbf{b}}^{(0)}$ where

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

From these limiting beliefs, using the initial beliefs $\tilde{\mathbf{b}}^{(0)}$, we can then calculate the *consensus* reached by all agents in the network. We have:

$$\frac{3}{7}\tilde{b}_1^{(0)} + \frac{2}{7}\tilde{b}_2^{(0)} + \frac{2}{7}\tilde{b}_3^{(0)} = \frac{3}{7} \frac{1}{3} + \frac{2}{7} \frac{1}{2} + 0 = 0.286$$

If the consensus is on the state {It is worth continuing studying}, this means that the three students reach a consensus for which they will agree that it is worth continuing studying with probability 0.286. This example shows that the beliefs converge over time for all the students and that they reach a consensus but it also shows that agent 1 has more influence than agents 2 and 3 over the limiting beliefs. Observe that, compared to the example in Appendix D where we don't differentiate between short-lived and long-lived friends, the consensus of continuing studying here leads to a lower probability since $0.286 < 0.333$, even though we update on the same matrix $\tilde{\mathbf{G}}^L = \mathbf{P}$. This is because of the influence of the short-lived friends who change the initial beliefs from $\mathbf{b}^{(0)}$ to $\tilde{\mathbf{b}}^{(0)} = \tilde{\mathbf{G}} \mathbf{b}^{(0)}$.

Along the lines of this model, a possible interpretation of our evidence is that the *strength of interactions* between two students may affect how much they learn, the human capital accumulation and how much they value achievement. It also shapes social norms that accumulate over time, which affect years of schooling both directly and indirectly. This idea is related to Akerlof's and Kranton's (2002) concept of identity in economics, where learning at school can be viewed within a process of identity formation, resource allocation, and social interaction. In other words, following the sociology literature, Akerlof and Kranton (2002) postulate that students often care less about their studies than about what their friends think.²¹

5.2 Additional evidence

Our analysis so far suggests that the distinction between long-lived and short-lived ties is important for understanding long-run peer effects in education. The mechanism for the

²¹This is also related to the empirical study of De Giorgi et al. (2010) which shows that students from Bocconi University in Italy are more likely to choose a major if many of their peers make the same choice. They also show that peers can divert students from majors in which they have a relative ability advantage, with adverse consequences on academic performance.

social effects is based on the idea that the convergence of preferences and the emergence of social norms among peers need long-term interactions. In this section, we aim at ruling out alternative explanations.

In our analysis, we identified long-lived ties as peers nominated in both Wave I and Wave II. This definition implies that long-lived ties are friends who are more likely to be peers at the time of college decisions. One could thus put forward another explanation of why friends at school may influence education decisions: it could be the timing of friendship or *decision proximity* so that friends in the last grades (grades 10 to 12) are likely to have an impact on college decision, regardless of whether these are long-lived or short-lived ties. In other words, is it really the frequency and strength of social interactions or is it the timing of friendship formation that is crucial for future educational outcomes? We would therefore like to disentangle between the *decision proximity* effect and the *strength of interaction* effect. For this purpose, we select students in the last grades (grades 10 to 12) and distinguish between short-lived and long-lived ties and examine the effect on college choices. We estimate a modified version of model (1), that is

$$\begin{aligned}
y_{i,r,t+1} = & \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t+1} + \phi^{S_1} \sum_{j=1}^{n_r} g_{ij,r,t}^{S_1} y_{j,r,t+1} + \phi^{S_2} \sum_{j=1}^{n_r} g_{ij,r,t}^{S_2} y_{j,r,t+1} + \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x'_{j,r,t,t+1} \gamma^L \\
& + \frac{1}{g_{i,r,t}^{S_1}} \sum_{j=1}^{n_r} g_{ij,r,t}^{S_1} x'_{j,r,t,t+1} \gamma^{S_1} + \mathbf{x}'_{i,r} \delta + \frac{1}{g_{i,r,t}^{S_2}} \sum_{j=1}^{n_r} g_{ij,r,t}^{S_2} x'_{j,r,t,t+1} \gamma^{S_2} + \eta_{r,t} + \epsilon_{i,r,t+1}.
\end{aligned}$$

We here distinguish between short-lived ties where best friends have only been nominated in Wave I (lower grades) and not in Wave II (later grades), i.e. $\phi^S = \phi^{S_1}$, from short-lived ties where best friends have only been nominated in Wave II and not in Wave I, i.e. $\phi^S = \phi^{S_2}$. If the decision proximity matters, then coefficient ϕ^{S_2} should be significant while ϕ^{S_1} should not.

Table 4 contains the estimation results. The empirical results reveal that the education decision of short-lived ties continues to show a non-significant effect on individual education outcomes, regardless of whether peers are interacting in lower or higher grades, highlighting the crucial role of long-lived ties in college decision.

Another concern is that peers nominated in different time periods may have a different long-run effect because students value peer characteristics differently in friendship decisions made over time. Do students select peers differently between the first and the second wave or is it really that distinct types of peers (short-lived versus long-lived ties) are of different importance? To disentangle these effects, we check whether students select peers differently

between the first and the second wave. Table 5 compares the observable characteristics of peers who only appear in Wave I, those who only appear in Wave II, and those who appear in both waves. One can see that, in fact, there are no differences between these peers in terms of observable characteristics. To further investigate this issue, we test whether link formation differs between different waves. For this purpose, we use model (6) and pool the data for Wave I ($t = 1$) and Wave II ($t = 2$):

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

In this model, $g_{ij,r,t} = 1$ if there is a link between i and j belonging to network r at time t (where $t = \text{Wave I, Wave II}$), $x_{i,r,t}^m$ indicates the individual characteristic m of individual i in network r at time t and $d_{ij,r}$ is a dummy variable, which is equal to 1 if a link $g_{ij,t}$ exists in Wave II, and zero otherwise. The parameter in front of the dummy variable captures the differences between the importance of these characteristics in link formation between Wave I and Wave II. Table 6 shows that most coefficients are not significant and that there are no observable differences in the link formation process between Waves I and II. We have also performed an F test that tests the joint significance of the γ parameters.²² Table 6 reports the p value of this test. It reveals that, controlling for network fixed effects, we cannot reject the null hypothesis of $\gamma^m = 0, \forall m = 1, \dots, M$. In summary, Tables 4 and 5 provide evidence showing that there are no differences between peers in Waves I and II in terms of observable characteristics and that the link formation between the different waves is not different. Finally, we investigate whether there are any structural differences across Wave I and Wave II in terms of the topology of the network. Over the past years, social network theorists have proposed a number of measures to account for the variability in network location across agents (Wasserman and Faust, 1994). We present those indicators in Appendix C where we define the density and the assortativity of a network and, at the node and network level, the betweenness centrality, the closeness centrality and the clustering coefficient. When applied to our Wave I and Wave II networks, we obtain the results collected in Table 7. It appears that the two networks are topologically very similar.

²²The idea is similar to the *Chow test* in time series analysis to investigate the existence of a structural break (see e.g. Chow, 1960; Hansen, 2000, 2001).

6 Robustness checks

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

6.1 Endogenous network formation

Our identification strategy hinges on the use of *network fixed effects* to control for unobserved factors driving both network formation and behavior over networks. The assumption is that such unobserved characteristics are common to the individuals within each network. As highlighted in Section 3.2, this is reasonable in our case study where the networks are quite small (see Section 2). However, if there are student-level unobservables that drive both network formation and outcome choice, this strategy fails. A possible way to tackle this issue is to *simultaneously* estimate network formation and outcomes. This strategy can be pursued by using parametric modeling assumptions and Bayesian inferential methods that allow to integrate a network formation with the study of behavior over the formed networks.²³

In this section, we develop this approach for our case and check whether our results still holds true when network formation is explicitly modeled.

Let us start with the network formation model. Under *homophily*, linked individuals are likely to be similar not only in terms of *observed characteristics* but also in terms of *unobserved characteristics* that could influence their behavior.²⁴ By failing to account for similarities in (unobserved) characteristics, similar behaviors might mistakenly be attributed to peer influence when they are simply due to similar characteristics. Let us consider a network formation model where the variables that explain the short- ($l = S$) or long-lived ($l = L$) tie between students i and j belonging to network r , i.e. $g_{ij,r}^l$, are the distances between them in terms of observed and unobserved characteristics:²⁵

$$P(g_{ij,r,t}^l = 1 | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) = \Lambda(\alpha^l + \sum_{m=1}^M \delta_m^l |x_{i,r}^m - x_{j,r}^m| + \theta^l |z_{i,r} - z_{j,r}| + \eta_r^l). \quad (6)$$

²³A similar approach, with a more articulated network formation model, has been used in Goldsmith-Pinkham and Imbens (2013).

²⁴This is also true under assortative matching (i.e. heterophily) but, in that case, there are links if people are dissimilar.

²⁵This is different to the network formation model in equation (5) since we explicitly model how the difference in observables affects link formation.

where $z_{i,r}$ denotes the *unobserved characteristic* at the individual level and $\Lambda(\cdot)$ is the logistic function. In (6), $P(g_{ij,r,t}^l = 1 | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r})$ is the probability that a link of type $l = S, L$ between students i and j is formed, given their observable and unobservable characteristics. Homophily behavior in the unobserved characteristics implies that $\theta^l < 0$, i.e. the closer two individuals are in terms of unobservable characteristics, the higher is the probability that they are friends. If $z_{i,r}$ is correlated with $\epsilon_{i,r}$, then the networks G_r^L and G_r^S in model (1) are endogenous. Following Hsieh and Lee (2015), we assume a bivariate normal distribution,

$$(z_{i,r}, \epsilon_{i,r}) \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \sigma_{\epsilon z} \\ \sigma_{\epsilon z} & \sigma_\epsilon^2 \end{pmatrix} \right). \quad (7)$$

It implies that $E(\epsilon_{i,r} | z_{i,r}) = \frac{\sigma_{\epsilon z}}{\sigma_z^2} z_{i,r}$, when conditioning on $z_{i,r}$. Hence, the outcome equation can be written as:

$$\begin{aligned} y_{i,r,t+1} = & \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t+1} + \phi^S \sum_{j=1}^{n_r} g_{ij,r,t+1}^S y_{j,r,t+1} + \mathbf{x}'_{i,r} \delta + \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x'_{j,r,t,t+1} \gamma^L \quad (8) \\ & + \frac{1}{g_{i,r,t}^S} \sum_{j=1}^{n_r} g_{ij,r,t}^S x'_{j,r,t,t+1} \gamma^S + \eta_{r,t} + \frac{\sigma_{\epsilon z}}{\sigma_z^2} z_{i,r} + u_{i,r,t+1}, \end{aligned}$$

where $u_{i,r,t+1} \sim N(0, \sigma_z^2 - \frac{\sigma_{\epsilon z}^2}{\sigma_z^2})$. Bayesian inference requires the computation of marginal distribution for all parameters. In particular, it requires the integration of complicated distributions, thus Markov Chain Monte Carlo (MCMC) techniques are usually employed to obtain random draws from posterior distributions. The unobservable variable ($z_{i,r}$) is thus generated according to the joint likelihood of link formation and outcome. The Gibbs sampling algorithm allows us to draw random values for each parameter from their posterior marginal distribution.²⁶

Table 8 panel (b) collects the results that are obtained when estimating simultaneously equations (1) and (6). Panel (a) shows the estimation result of the model, without distinguishing between short-lived and long lived ties (homogeneous peer effects). The first column in both panels reports the 2SLS results for comparison. Table 8 reveals that $\sigma_{\epsilon z}$ is not significantly different from zero for both the models with homogeneous and heterogeneous peer effects (columns (3) and (6) of Table 8). The Bayesian estimates are close to the

²⁶See Appendix E for more details on the estimation procedure. An introduction to Monte Carlo methods in Bayesian econometrics can be found in Chib (1996) and Casella and Robert (2004).

2SLS estimates.²⁷ This evidence is thus in support of our identification strategy. Indeed, our list of controls and network fixed effects, together with the temporal lag between when friends are chosen and when education levels are attained, seem to account for unobserved factors driving both network formation and behavior over networks.

6.2 Measurement errors in network links

In this section, we perform two different robustness checks.

6.2.1 Directed networks

First, our empirical investigation has assumed that friendship relationships are symmetric, i.e. $g_{ij} = g_{ji}$. We check here how sensitive our results are to such an assumption, i.e. to a possible measurement error in the definition of the peer group. Indeed, our data make it possible to know exactly who nominates whom in a network and we find that 12 percent of the relationships in our dataset are not reciprocal. Instead of constructing an undirected network, in this section, we perform our analysis using *directed networks*. We focus on the choices made (outdegrees) and we denote a link from i to j as $g_{ij,r} = 1$ if i has nominated j as his/her friend in network r , and $g_{ij,r} = 0$, otherwise.²⁸ Table 9 shows the estimation results of model (1) for directed networks. The results remain qualitatively unchanged and only slightly higher in magnitude.

6.2.2 A simulation experiment

Second, our identification and estimation strategies depend on the correct identification of long-lived and short-lived ties. In this section, we test the robustness of our results with respect to misspecification of long-lived and short-lived network topologies. Indeed, in our theoretical model, we assume that $\phi^L > \phi^S$ and our empirical analysis confirms this assumption by finding a significant effect of long-lived ties (but not short-lived ties) on educational outcomes. These results clearly depend on the definition of a long-lived and a short-lived tie. In the present robustness check, we want to check whether our results are

²⁷Those estimates are slightly different from those collected in Table 3 because the computational burden of Bayesian estimation forced us to drop small networks, since they create computational problems when some covariates are constant across the same network, and big networks, because they slow the computation time excessively.

²⁸As highlighted by Wasserman and Faust (1994), centrality indices for directional relationships generally focus on choices made (outdegrees). The estimation results, however, remain qualitatively unchanged if we define the link using the nominations received (indegrees).

robust even if we fail in exactly identifying long-lived and short-lived ties. To be more precise, we use simulated data to answer such questions as: Do our results change if some links are not assigned in the right category (short-lived or long-lived ties)? Do our results change if some links are not reported? To what extent? How many ties need to be misspecified before our results disappear?

In our analysis, we have defined a long-lived tie as a friend nominated twice, and a short-lived tie as a friend nominated just once. We can imagine that a student may be more likely to report a long-lived tie than a short-lived tie. Let us suppose that individual i reports a long-lived tie ($l = L$) with probability p , a short-lived tie ($l = S$) with probability q , with $q < p$, and another individual (neither a short- nor a long-lived tie, $l = N$) with probability r , with $r < q < p$.

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

| | | Observed | | | |
|------|---|----------|-----------|-----------|---|
| | | L | S | N | |
| True | L | p^2 | $2p(1-p)$ | $(1-p)^2$ | 1 |
| | S | q^2 | $2q(1-q)$ | $(1-q)^2$ | 1 |
| | N | r^2 | $2r(1-r)$ | $(1-r)^2$ | 1 |
| | | s | w | n | |

For example, a long-lived tie appears as a long-lived tie with probability p^2 , as a short-lived tie with probability $2p(1-p)$, and may be missed with probability $(1-p)^2$. In the table, $s = p^2 + q^2 + r^2$ denotes the probability of observing a long-lived tie, $w = 2p(1-p) + 2q(1-q) + 2r(1-r)$ denotes the probability of observing a short-lived tie and $n = (1-p)^2 + (1-q)^2 + (1-r)^2$ is the probability of not observing a tie.

Our empirical analysis assumes $s = p^2$, $w = 2q(1-q)$, $n = (1-r)^2$ and that the off diagonal elements are equal to zero. A misspecification of the network topology implies that the off diagonal elements are different from zero. Let us denote these off diagonal elements as P_{LM} , which are the probabilities of moving from state L to state M ; $L, M = \{S, L\}$. In our numerical exercise, we gradually change those elements from 0 to 1 at a pace of 0.005, i.e. $P_{LM} = [0, 0.005, 0.010, \dots, 1]$.

Our misspecification experiment can be summarized by the following table:

| | | Observed | | |
|------|---|----------|----------|----------|
| | | L | S | N |
| True | L | · | P_{SW} | P_{SN} |
| | S | P_{WS} | · | P_{WN} |
| | N | P_{NS} | P_{NW} | · |

For ease of computation, we proceed in two steps. First, we change ties from long-lived to short-lived and vice versa, i.e. we change $P_{LS} = \frac{2p(1-p)}{p^2+2p(1-p)}$ and $P_{SL} = \frac{q^2}{q^2+2q(1-q)}$. Second, for each combination of P_{LS} and P_{SL} , we change ties to non-ties and vice versa, i.e. we change $P_{LN} = \frac{(1-p)^2}{p^2+2p(1-p)+(1-p)^2}$, $P_{NL} = \frac{n^2}{n^2+2n(1-n)+(1-n)^2}$, $P_{SN} = \frac{(1-q)^2}{q^2+2q(1-q)+(1-q)^2}$ and $P_{NS} = \frac{(1-n)^2}{n^2+2n(1-n)+(1-n)^2}$.

In this framework, the higher are these probabilities, the further away we are from our observed network topology. For example, the combination $P_{LS} = 0.1$ and $P_{LN} = 0$ means that 10% of the long-lived ties are replaced by short-lived ties; the combination $P_{LS} = 0.3$ and $P_{LN} = 0.2$ means that 30% of the long-lived ties are replaced by short-lived ties and 20% of the short-lived ties are replaced by unconnected individuals. In other words, our experiment does not only allow for the fact that long-lived and short-lived ties are not equally likely to be interchanged, but also considers the possibility that they each have some probability of generating a misreport that violates the exclusion restrictions. For each combination of P_{LS} , P_{SL} , P_{LN} , P_{NL} , P_{SN} and P_{SW} , we draw one hundred network structures (samples) of a size equal to the real one ($n = 1,819$). Then, we estimate model (3) replacing the real G_r^L and G_r^S matrices with the simulated ones in turn so that, in total, we estimate model (3) eighty thousand times for each type of estimator described in Appendix B.

Note that this exercise is quite similar to directly changing p , q and r . The advantage of our approach is that it does not need to specify p , q and r . Indeed, p , q and r are not known by the econometrician. They can be estimated imposing that observed and true numerosity are the same for each type of tie, but there is not any clear theoretical reason why this should be the case. An exploration of the entire space spanned by (p, q, r) would imply a change in the observed (or true) network density which, in turn, would render our peer effect estimates non-comparable among combinations.

Simulated evidence

Figure 1 shows the results of our simulation experiment for the *2SLS bias-corrected lagged estimator*.²⁹ It depicts the estimates of long-lived and short-lived tie effects with 90% con-

²⁹The simulation results for the other estimators are similar. They are available upon request.

fidence bands, in the upper and lower panel, respectively.³⁰ The first important question concerns the percentage of network-structure misspecifications needed for the long-lived tie effects on college choice to disappear. The upper panel of Figure 1 shows the estimates for each combination of replacement rates – between long-lived and short-lived ties (P_{LS}) and between long-lived ties and no ties (P_{LN}). The graph shows that long-lived tie effects remain statistically significant for levels of P_{LS} and P_{LN} in the range of 0.005 and 0.35. Figure 2 depicts the conditional results (i.e. P_{LS} conditional on $P_{LN} = 0$ in the upper panel and P_{LN} conditional on $P_{LS} = 0$ in the lower panel). The upper panel shows that long-lived-tie effects remain statistically significant up to a percentage of randomly replaced links with short-lived ties of about 35%. The lower panel shows a similar result when increasing the percentage of links randomly replaced by zeros. This evidence implies that even if we do not observe or we imprecisely observe a portion of each individual’s long-lived ties, our results on the existence of this effect still hold. The second question is what is the percentages of replacement needed in order to have a significant effect of short-lived ties. The lower panel of Figure 1 shows the estimates of short-lived tie effects for each combination of replacement rates – between short-lived and long-lived ties (P_{SL}) and between short-lived ties and no ties (P_{SN}). The graph shows that we need to replace almost 70% of the short-lived ties with long-lived ties before finding an effect which is statistically different from 0. Naturally, when replacing short-lived ties with no ties, we continue to detect no effect and the standard error increases with the percentage of replaced links. The lower panel of Figure 2 shows this evidence more clearly by depicting the conditional results (i.e. P_{SL} conditional on $P_{SN} = 0$ in the upper panel and P_{SN} conditional on $P_{SL} = 0$ in the lower panel). These results show that the effects of short-lived ties are found to be important only when the large majority of long-lived ties is labeled as short-lived ties.

Finally, we show in Figure 3 the rejection rates³¹ when using the *2SLS bias-corrected estimator* and the *2SLS bias-corrected lagged estimator*. This graph indicates that the *2SLS bias-corrected lagged estimator* tends to be more robust to a possible misspecification of

³⁰Standard errors have been calculated assuming drawing independence and taking into account the variation between estimates for each replacement rate. Specifically, the standard error at each replacement rate, say i , is computed as follows:

$$\sigma_i = \sqrt{W_i + B_i}$$

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

³¹Rejection refers to the null hypothesis of having $\phi^L = 0$ or $\phi^S = 0$, respectively, for long-lived and short-lived ties effects. Each rate represents the frequency of rejection for the corresponding percentage of randomly replaced links.

long-lived and short-lived ties. Indeed, it appears that this estimator needs, on average, a higher percentage of misspecified ties to accept the hypothesis of no effects for long-lived ties and to reject it for short-lived ties. To wrap up, in this section, we have shown that the strength of long-lived ties ϕ^L is reduced and becomes insignificant when we have converted more than 35% of the long-lived ties into short-lived ties while the strength of short-lived ties ϕ^S is increasing and becomes significant after having replaced more than 60% of the short-lived ties with long-lived ties. To illustrate this result, consider a student i who has twenty friends, ten long-lived ties $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and ten short-lived ties $\{11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$. Even if we incorrectly assign three friends from one category (long-lived tie) to the other (short-lived tie), our results will still hold. For instance, if we instead observe $\{11, 12, 3, 4, 5, 6, 7, 8, 9, 10\}$ as long-lived ties (labeling 11 and 12 as long-lived when they are short-lived ties) and $\{1, 2, 13, 14, 15, 16, 17, 18, 19, 20\}$ as short-lived ties (labeling 1 and 2 as short-lived when they are long-lived ties), we would still have a significant effect of long-lived ties on education and a non-significant effect of short-lived ties since we have “only” converted 30% of the links. As a result, from 3 to 8 incorrect assignments (which correspond to 30% to 80% conversion of long-lived ties into short-lived ties or the contrary), both effects will still be insignificant. It is only after having converted seven out of ten ties (i.e. more than 60% of the long-lived ties have been converted into short-lived ties, or the contrary) that we find that short-lived ties have a significant effect on education while long-lived ties do not.

7 Short-run versus long-run effects

So far, we have found that students nominate other students as their best friends but only their long-lived ties (i.e. students who are friends in both waves) influence them in their educational choices. Using Addhealth data for Wave I only, Calvó-Armengol et al. (2009) have studied the *current* effect of peers on education, finding that peers do affect the current education activity (i.e. grades) of students. They did not differentiate between different types of peers.

To further investigate this issue, we would now like to oppose the long-run effects to the short-run effects of peers on education by differentiating between the effect of long-lived ties and short-lived ties on school performance. For this purpose, we estimate the short-run

counterpart of equation (1):

$$\begin{aligned}
y_{i,r,t} = & \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t} + \phi^S \sum_{j=1}^{n_r} g_{ij,r,t}^S y_{j,r,t} + \mathbf{x}'_{i,r} \delta \\
& + \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x'_{j,r,t} \gamma^L + \frac{1}{g_{i,r,t}^S} \sum_{j=1}^{n_r} g_{ij,r,t}^S x'_{j,r,t} \gamma^S + \eta_{r,t} + \epsilon_{i,r,t},
\end{aligned} \tag{9}$$

where $y_{i,r,t}$ is now the grade of student i who belongs to network r at time t where t refers to Wave II. The rest of the notation remains unchanged, which implies that we now deal with a traditional peer effects model where all individual and peer group characteristics are contemporaneous (i.e. in Wave II in 1995-1996). As in our investigation on long-run effects, we exploit variations in link formation in Waves I and II to differentiate between long-lived ties and short-lived ties. We then look at how these different types of peers affect each student's grade obtained in Wave II. The identification and estimation strategy remains unchanged with the difference that now, we cannot use IV variables lagged in time (see Appendix B).

School performance is measured using the respondent's scores received in Wave II in several subjects, namely English or language arts, history or social science, mathematics and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. For each individual, we calculate an index of school performance using a standard principal component analysis. The final composite index (labeled as GPA index or grade point average index) is the first principal component.³² It ranges between 0 and 6.09, with a mean equal to 2.29 and a standard deviation equal to 1.49.

The estimation results of model (9) are contained in Table 10. It appears that *while in the long run, only long-lived ties matter, in the short run, both short and long-lived ties are important in determining a student's performance at school*. A standard deviation increase in aggregate GPA of peers translates, respectively, into a 8.1 (for long-lived ties) and a 4.8 (for short-lived ties) percent increase of a standard deviation in the individual's GPA. Taking our analysis as a whole, our results suggest that, in the short run, all peers matter for education (i.e. grades) while, in the long run, only long-lived ties matter for future educational choices (i.e. years of schooling). This is consistent with the model developed in Section 5.1 where, in the very short-run, both long-lived and short-lived friends have an impact on current beliefs while, in the long run, only long-lived friends have an impact on educational decisions since

³²The index explains roughly 56 percent of the total variance and captures a general performance at school since it is positively and highly correlated with the scores in all subjects. Further details on this procedure are available upon request.

the latter are influenced by the social norm emerging from the iterations of beliefs. This is also consistent with the fact that, in the short-run analysis, the outcome is the students' grades while, in the long-term, the outcome is the number years of education, where social norms matter more.

8 Concluding remarks

In this paper, we consider the estimation of heterogeneous spillover effects in a network model by looking at the impact of different types of friends made at school on education decisions. We find that a *long-lived* relationship has a positive impact on own education outcomes while a *short-lived* relationship does not. We present a theoretical model providing a mechanism that is consistent with our evidence. In our model, short-lived friends affect each other only in the initial period while long-lived friends affect each other all the time. We show how a social norm that favors education arises in the long run and that only long-term relationships between peers have a crucial influence on the emergence of this norm. Our results provide new evidences about peer effects heterogeneity in achievement, which may be an issue when the policy maker wants to determine the effects of regrouping students (Fruehwirth, 2014). Such policy tool manipulates the length of interactions among students, which, as we show in this paper, produces different peer effects. Consequently our evidences highlight that further research needs to be done to robustly understand the effects of regrouping.

In line with several studies in sociology and economics (e.g. Coleman, 1988, Wellman and Wortley, 1990, Akerlof and Kranton, 2002), our results suggest that long-lasting social interactions affect how much students value achievement, their human capital accumulation, and how social norms are formed. On the other hand, we find that, in the short run, any relationship (whether it is a long or short lived) has an impact on current grades. Our analysis suggests that more research is needed to further understand the complex relationships between different types of peers and how they affect educational outcomes.

References

- [1] Akerlof, G.A. and R.E. Kranton (2002), “Identity and schooling: Some lessons for the economics of education,” *Journal of Economic Literature* 40, 1167-1201.
- [2] Bekker, P. (1994), “Alternative approximations to the distributions of instrumental variable estimators,” *Econometrica* 62, 657-681.
- [3] Bekker, P. and J. van der Ploeg (2005), “Instrumental variable estimation based on grouped data,” *Statistica Neerlandica* 59, 239-267.
- [4] Bifulco, R., Fletcher, J.M. and S.L. Ross (2011), “The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health,” *American Economic Journal: Economic Policy* 3, 25-53.
- [5] Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), “Identification of social interactions,” In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 853-964.
- [6] Boucher, V. and B. Fortin (2015), “Some challenges in the empirics of the effects of networks,” In: Y. Bramoullé, B.W. Rogers and A. Galeotti (Eds.), *Oxford Handbook on the Economics of Networks*, Oxford: Oxford University Press, forthcoming.
- [7] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), “Identification of peer effects through social networks,” *Journal of Econometrics* 150, 41-55.
- [8] Brooks-Gunn, J., Duncan, G.J., Kato Klebanov, P. and N. Sealant (1993), “Do neighborhoods influence child and adolescent development,” *American Journal of Sociology* 99, 353-395.
- [9] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), “Peer effects and social networks in education,” *Review of Economic Studies* 76, 1239-1267.
- [10] Casella, G. and C. Robert (2004), *Monte Carlo Statistical Methods*, Berlin: Springer Verlag.
- [11] Chib, S. (1996), “Calculating posterior distributions and modal estimates in Markov mixture models,” *Journal of Econometrics* 75, 79-97.
- [12] Chow, G.C. (1960), “Test of equality between sets of coefficients in two linear regressions,” *Econometrica* 28, 591-603.

- [13] Coleman, J. (1988), "Social capital in the creation of human capital," *American Journal of Sociology* 94, 95-120.
- [14] Corcoran, M., Gordon, R., Laren, D. and G. Solon (1992), "The association between men's economic status and their family and community origins" *Journal of Human Resources* 27, 575-601.
- [15] Datcher, L. (1982), "Effects of community and family background on achievement," *Review of Economics and Statistics* 64, 32-41.
- [16] DeGroot, M. H. (1974), "Reaching a consensus," *Journal of the American Statistical Association* 69, 118-121.
- [17] De Giorgi, G., Pellizzari, M. and S. Redaelli (2010), "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics* 2, 241-275.
- [18] DeMarzo, P., D. Vayanos, and J. Zwiebel (2003), "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics* 118, 909-968.
- [19] Donald, S.G., and W.K. Newey (2001), "Choosing the number of instruments," *Econometrica* 1161-1191.
- [20] Durlauf, S.E. (2004), "Neighborhood effects," In: J.V. Henderson and J-F. Thisse (Eds.), *Handbook of Regional and Urban Economics Vol. 4*, Amsterdam: Elsevier Science, pp. 2173-2242.
- [21] Evans, W.N, Oates, W.E and R.M. Schwab (1992), "Measuring peer group effects: A study of teenage behavior," *Journal of Political Economy* 100, 966-991.
- [22] Fruehwirth, J. C. (2014), "Can achievement peer effect estimates inform policy? A view from inside the black box, " *Review of Economics and Statistics* 96(3), 514-523.
- [23] Goldsmith-Pinkham, P. and G.W. Imbens (2013), "Social networks and the identification of peer effects," *Journal of Business and Economic Statistics* 31, 253-264.
- [24] Golub, B. and M.O. Jackson (2010), "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics* 2, 112-149.
- [25] Golub, B. and M.O. Jackson (2012), "How homophily affects the speed of learning and best-response dynamics," *Quarterly Journal of Economics* 127, 1287-1338.

- [26] Graham, B.S. (2015), “Methods of identification in social networks,” *Annual Review of Economics*, forthcoming.
- [27] Granovetter, M.S. (1973), “The strength of short-lived ties,” *American Journal of Sociology* 78, 1360-1380.
- [28] Granovetter, M.S. (1974), *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.
- [29] Granovetter, M.S. (1983), “The strength of short-lived ties: A network theory revisited,” *Sociological Theory* 1, 201-233.
- [30] Griffith, A.L. and K.N. Rask (2014), “Peer effects in higher education: A look at heterogeneous impacts,” *Economics of Education Review* 39, 65-77.
- [31] Hansen, B.E. (2000), “Testing for structural change in conditional models,” *Journal of Econometrics* 97, 93-115.
- [32] Hansen, B.E. (2001), “The new econometrics of structural change: Dating breaks in U.S. labor productivity,” *Journal of Economic Perspectives* 15, 117-128.
- [33] Hansen, C., Hausman, J. and W. Newey (2008), “Estimation with many instrumental variables,” *Journal of Business and Economic Statistics* 26, 398-422.
- [34] Hsieh, C.-S. and L.-F. Lee (2015), “A social interaction model with endogenous friendship formation and selectivity,” *Journal of Applied Econometrics*, forthcoming.
- [35] Ioannides, Y.M. (2011), “Neighborhood effects and housing,” In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 1281-1340.
- [36] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.
- [37] Ioannides, Y.M. and G. Topa (2010), “Neighborhood effects: Accomplishments and looking beyond them,” *Journal of Regional Science* 50, 343-362.
- [38] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.

- [39] Jackson, M.O. and Y. Zenou (2015), “Games on networks,” In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, Amsterdam: Elsevier, pp. 91-157.
- [40] Lai, G., N. Lin and S.-Y. Leung (1998), “Network resources, contact resources, and status attainment,” *Social Networks* 20, 159-178.
- [41] Lee, L-F. (2002), “Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models,” *Econometric Theory* 18, 252-277.
- [42] Lee, L-F. and Liu X. (2010), “Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances,” *Econometric Theory* 26, 187-230.
- [43] Lee, L-F., Liu, X. and X. Lin (2010), “Specification and estimation of social interaction models with network structures,” *Econometrics Journal* 13, 145-176.
- [44] Lin, X. (2010), “Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables,” *Journal of Labor Economics* 28, 825-860.
- [45] Lin, N., W.M. Ensel and J.C. Vaughn (1981), “Social resources and strength of ties: Structural factors in occupational status attainment,” *American Sociological Review* 46, 393-405.
- [46] Liu, X. and L-F. Lee (2010), “GMM estimation of social interaction models with centrality,” *Journal of Econometrics* 159, 99-115.
- [47] Liu, X., Patacchini, E., Zenou, Y. and L-F. Lee (2012), “Criminal networks: Who is the key player?” CEPR Discussion Paper No. 8772.
- [48] Manski, C.F. (1993), “Identification of endogenous effects: The reflection problem,” *Review of Economic Studies* 60, 531-542.
- [49] Marsden, P.V. and J.S. Hurlbert (1988), “Social resources and mobility outcomes: A replication and extension,” *Social Forces* 66, 1038-1059.
- [50] Patacchini, E. and Y. Zenou (2008), “The strength of weak ties in crime,” *European Economic Review* 52, 209-236.
- [51] Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.

- [52] Sacerdote, B. (2011), “Peer effects in education: How might they work, how big are they and how much do we know thus far?”, In: E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of Economics of Education, Vol. 3*, Amsterdam: Elsevier Science, pp. 249-277.
- [53] Tincani, M.M. (2015), “Heterogeneous peer effects and rank concerns: Theory and evidence,” Unpublished manuscript, University College London.
- [54] Wasserman, S. and K. Faust (1994), *Social Network Analysis. Methods and Applications*, Cambridge: Cambridge University Press.
- [55] Wellman, B. and S. Wortley (1990), “Different strokes from different folks: Community ties and social support,” *American Journal of Sociology* 96, 558-588.
- [56] Yakubovich, V. (2005), “Short-lived ties, information, and influence: How workers find jobs in a local Russian labor market,” *American Sociological Review* 70, 3, 408-421.
- [57] Yakusheva, O., Kapinos, K.A. and D. Eisenberg (2014), “Estimating heterogeneous and hierarchical peer effects on body weight using roommate assignments as a natural experiment,” *Journal of Human Resources* 49, 234-261.
- [58] Zax, J.S. and D.I. Rees (2002), “IQ, academic performance, environment, and earnings,” *Review of Economics and Statistics* 84, 600-616.

Appendix A: Data Appendix

Table A1 provides a detailed description of the variables used in our study as well as the summary statistics for our sample. Among the individuals selected in our sample, 53 percent are female and 19 percent are blacks. The average parental education is high-school graduate. Roughly 10 percent have parents working in a managerial occupation, another 10 percent in the office or sales sector, 20 percent in a professional/technical occupation, and roughly 30 percent have parents in manual occupations. More than 70 percent of our individuals come from households with two married parents and from households of about four people on average. In Wave IV, 42 percent of our adolescents are now married and nearly half of them (43 percent) have at least a son or a daughter. The mean intensity in religion practice slightly decreases during the transition from adolescence to adulthood. On average, during their teenage years, our individuals felt that adults care about them and they had a good relationship with their teachers. Roughly, 30 percent of our adolescents were high-performing individuals at school, i.e. had the highest mark in mathematics. On average, these adolescents declare having the same number of best friends both in Wave I and Wave II (about 2.50 friends), although the composition of the friends changes.

Appendix B: Estimation - Technical details -

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

$$\mathbf{Y} = \mathbf{Z}\theta + \iota \cdot \eta + \epsilon, \quad (10)$$

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

We treat η as a vector of unknown parameters. When the number of networks \bar{r} is large, we have the incidental parameter problem. Let $\mathbf{J} = D(\mathbf{J}_1, \dots, \mathbf{J}_{\bar{r}})$, where $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r} \mathbf{1}'_{n_r} \mathbf{1}_{n_r}$. The network fixed effect can be eliminated by a transformation with \mathbf{J} such that:

$$\mathbf{JY} = \mathbf{JZ}\theta + \mathbf{J}\epsilon. \quad (11)$$

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

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

It follows that $\mathbf{G}^L \mathbf{Y} = \mathbf{G}^L \mathbf{M} \mathbf{X}^* \beta + \mathbf{G}^L \mathbf{M} \iota \eta + \mathbf{G}^L \mathbf{M} \epsilon$ and $\mathbf{G}^S \mathbf{Y} = \mathbf{G}^S \mathbf{M} \mathbf{X}^* \beta + \mathbf{G}^S \mathbf{M} \iota \eta + \mathbf{G}^S \mathbf{M} \epsilon$. $\mathbf{G}^L \mathbf{Y}$ and $\mathbf{G}^S \mathbf{Y}$ are correlated with ϵ because $E[(\mathbf{G}^L \mathbf{M} \epsilon)' \epsilon] = \sigma^2 \text{tr}(\mathbf{G}^L \mathbf{M}) \neq 0$ and $E[(\mathbf{G}^S \mathbf{M} \epsilon)' \epsilon] = \sigma^2 \text{tr}(\mathbf{G}^S \mathbf{M}) \neq 0$. Hence, in general, (11) cannot be consistently estimated by OLS.³³ If \mathbf{G} is row-normalized such that $\mathbf{G} \cdot \mathbf{l}_n = \mathbf{l}_n$, where \mathbf{l}_n is a n -dimensional vector of ones, the endogenous social interaction effect can be interpreted as an average effect. Liu and Lee (2010) use an instrumental variable approach and propose different estimators based on different instrumental matrices, here denoted by \mathbf{Q}_1 and \mathbf{Q}_2 . They first consider the 2SLS estimator based on the conventional instrumental matrix for the estimation of (11): $\mathbf{Q}_1 = \mathbf{J}(\mathbf{G} \mathbf{X}^*, \mathbf{X}^*)$ (*finite-IVs 2SLS*). Then, they propose to use additional instruments (IVs) $JG\iota$ and enlarge the instrumental matrix: $\mathbf{Q}_2 = (\mathbf{Q}_1, JG\iota)$ (*many-IVs 2SLS*). The additional IVs of $JG\iota$ are based on the row sums of G and are indicators of centrality in the networks. Liu and Lee (2010) show that those additional IVs could help model identification when the conventional IVs are short-lived and improve on the estimation efficiency of the conventional 2SLS estimator based on \mathbf{Q}_1 . However, the number of such instruments depends on the number of networks. If the number of networks grows with the sample size, so does the

³³Lee (2002) has shown that the OLS estimator can be consistent in the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small. However, in the current data, the number of neighbors are limited, and hence that result does not apply.

number of IVs. The 2SLS could be asymptotic biased when the number of IVs increases too fast relative to the sample size (see, e.g., Bekker, 1994; Bekker and van der Ploeg, 2005; Hansen et al., 2008). As detailed in Section 2, in this empirical study, we have a number of small networks. Liu and Lee (2010) also propose a bias-correction procedure based on the estimated leading-order many-IV bias (*bias-corrected 2SLS*). The bias-corrected many-IV 2SLS estimator is properly centered, asymptotically normally distributed, and efficient when the average group size is sufficiently large. Thus, it is the more appropriate estimator in our case study.³⁴

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

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

which is an $n \times (3m + 2)$ matrix. However, this matrix is infeasible as it involves unknown parameters. Note that f can be considered as a linear combination of the vectors in $\mathbf{Q}_0 = J[\mathbf{G}^L \mathbf{M}(\mathbf{X}^* + \iota), \mathbf{G}^S \mathbf{M}(\mathbf{X}^* + \iota), \mathbf{X}^*]$. As ι has \bar{r} columns the number of IVs in \mathbf{Q}_0 increases as the number of groups increases. Furthermore, as $\mathbf{M} = (\mathbf{I} - \phi^L \mathbf{G}^L - \phi^S \mathbf{G}^S)^{-1} = \sum_{j=0}^{\infty} (\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S)^j$ when $\sup \|\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\|_{\infty} < 1$, $\mathbf{M}\mathbf{X}^*$ and $\mathbf{M}\iota$, can be approximated by linear combinations of

$$(\mathbf{G}^L \mathbf{X}^*, \mathbf{G}^S \mathbf{X}^*, \mathbf{G}^S \mathbf{G}^L \mathbf{X}^*, (\mathbf{G}^L)^2 \mathbf{X}^*, (\mathbf{G}^S)^2 \mathbf{X}^*, (\mathbf{G}^S)^2 \mathbf{G}^L \mathbf{X}^*, (\mathbf{G}^S)^2 (\mathbf{G}^L)^2 \mathbf{X}^*, \dots)$$

and

$$(\mathbf{G}^L \iota, \mathbf{G}^S \iota, \mathbf{G}^S \mathbf{G}^L \iota, (\mathbf{G}^L)^2 \iota, (\mathbf{G}^S)^2 \iota, (\mathbf{G}^S)^2 \mathbf{G}^L \iota, (\mathbf{G}^S)^2 (\mathbf{G}^L)^2 \iota, \dots),$$

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

$$\begin{aligned} \mathbf{Q}_{\infty} = & \mathbf{J}(\mathbf{G}^L(\mathbf{G}^L \mathbf{X}^*, \mathbf{G}^S \mathbf{X}^*, \mathbf{G}^S \mathbf{G}^L \mathbf{X}^*, \dots, \mathbf{G}^L \iota, \mathbf{G}^S \iota, \mathbf{G}^S \mathbf{G}^L \iota, \dots), \\ & \mathbf{G}^S(\mathbf{G}^L \mathbf{X}^*, \mathbf{G}^S \mathbf{X}^*, \mathbf{G}^S \mathbf{G}^L \mathbf{X}^*, \dots, \mathbf{G}^L \iota, \mathbf{G}^S \iota, \mathbf{G}^S \mathbf{G}^L \iota, \dots), \mathbf{X}^*). \end{aligned}$$

Let $\mathbf{Q}_{\mathbf{K}}$ be an $n \times K$ submatrix of \mathbf{Q}_{∞} (with $K \geq 3m + 2$) including \mathbf{X}^* . Let Q_S be an $n \times K_S$ submatrix of $Q_{S\infty} = \mathbf{G}^L(\mathbf{G}^L \mathbf{X}^*, \mathbf{G}^S \mathbf{X}^*, \mathbf{G}^S \mathbf{G}^L \mathbf{X}^*, \dots, \mathbf{G}^L \iota, \mathbf{G}^S \iota, \mathbf{G}^S \mathbf{G}^L \iota, \dots)$ and Q_S an $n \times K_S$ submatrix of $Q_{W\infty} = \mathbf{G}^S(\mathbf{G}^L \mathbf{X}^*, \mathbf{G}^S \mathbf{X}^*, \mathbf{G}^S \mathbf{G}^L \mathbf{X}^*, \dots, \mathbf{G}^L \iota, \mathbf{G}^S \iota, \mathbf{G}^S \mathbf{G}^L \iota, \dots)$. We assume that $\frac{K_S}{K} = 1$. Let $\mathbf{P}_{\mathbf{K}} = \mathbf{Q}_{\mathbf{K}}(\mathbf{Q}_{\mathbf{K}}' \mathbf{Q}_{\mathbf{K}})^{-1} \mathbf{Q}_{\mathbf{K}}'$ be the projector of $\mathbf{Q}_{\mathbf{K}}$. The resulting

³⁴Liu and Lee (2010) also generalize this 2SLS approach to the GMM using additional quadratic moment conditions.

2SLS estimator is given by:

$$\widehat{\theta}_{2sls} = (\mathbf{Z}'\mathbf{P}_K\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_K\mathbf{y}. \quad (14)$$

Asymptotic properties of 2SLS estimator

As shown by Liu and Lee (2010), the 2SLS with a fixed number of IVs would be consistent but not efficient. Asymptotic efficiency can be achieved using a sequence of IVs in which the number of IVs grows slow enough relative to the sample size. In general, K may be seen as an increasing function of n . Following Liu and Lee (2010), we assume the following regularity conditions:

Assumption C1: The elements of ϵ are i.i.d with zero mean, variance σ^2 and a moment of order higher than four exists.

Assumption C2: The elements of \mathbf{X}^* are uniformly bounded constants, \mathbf{X}^* has the full rank k and $\lim_{n \rightarrow \infty} \mathbf{X}^{*\prime}\mathbf{X}^*$ exists and is nonsingular.

Assumption C3: The sequences of matrices $\{\mathbf{G}^L\}$, $\{\mathbf{G}^S\}$, $\{\mathbf{M}\}$ are uniformly bounded.

Assumption C4: $\overline{\mathbf{H}} = \lim_{n \rightarrow \infty} \frac{1}{n}f'f$ is a finite non singular matrix.

Assumption C5: There exists a $\mathbf{K} \times (3m+2)$ matrix π_K such that $\|f - \mathbf{Q}_K\pi_K\|_\infty \rightarrow 0$ as $n, K \rightarrow \infty$.

The 2SLS estimator with an increasing number of IVs approximating f can be asymptotically efficient under some conditions. However, when the number of instruments increases too fast, such an estimator could be asymptotically biased, which is known as the many-instrument problem. Let $\Psi_{K,S} = \mathbf{P}_K\mathbf{G}^L\mathbf{M}$ and $\Psi_{K,W} = \mathbf{P}_K\mathbf{G}^S\mathbf{M}$. The following proposition shows consistency and asymptotic normality of the 2SLS estimator (14).

Under assumptions C1-C5, if $K/n \rightarrow 0$, then $\sqrt{n}(\widehat{\theta} - \theta - b_{2sls}) \xrightarrow{d} N(0, \sigma^2\overline{\mathbf{H}}^{-1})$, where $b_{2sls} = \sigma^2(\mathbf{Z}'\mathbf{P}_K\mathbf{Z})^{-1}[\text{tr}(\Psi_{K,L}), \text{tr}(\Psi_{K,S}), \mathbf{0}_{3m \times 1}]' = O_p(K/n)$.

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

$$\begin{aligned} \frac{1}{n}\mathbf{Z}'\mathbf{P}_K\mathbf{Z} &= \mathbf{H} - e_f + \frac{1}{n}v'\mathbf{P}_Kf + \frac{1}{n}f'\mathbf{P}_Kv + \frac{1}{n}v'\mathbf{P}_Kv \\ &= \mathbf{H} + \mathbf{O}(\text{tr}(e_f)) + \mathbf{O}_p(\sqrt{K/n}) + \mathbf{O}_p(K/n) \\ &= \overline{\mathbf{H}} + o_p(1) \end{aligned}$$

where $\mathbf{H} = \frac{1}{n}f'f$ and $e_f = \frac{1}{n}f'(I - \mathbf{P}_K)f$, because $e_f = \mathbf{O}(\text{tr}(e_f))$, $\frac{1}{n}v'\mathbf{P}_Kv = \mathbf{O}_p(K/n)$ and

$\frac{1}{n}v'P_{\mathbf{K}}f = O_p(\sqrt{K/n})$. Furthermore, we have

$$\begin{aligned}
& (\mathbf{Z}'P_{\mathbf{K}}\epsilon - \sigma^2 [\text{tr}(\Psi_{K,L}), \text{tr}(\Psi_{K,S}), \mathbf{0}_{3m \times 1}]') / \sqrt{n} \\
= & h - f'(I - P_{\mathbf{K}})\epsilon / \sqrt{n} + \left(\frac{1}{n}v'P_{\mathbf{K}}\epsilon - \sigma^2 [\text{tr}(\Psi_{K,L}), \text{tr}(\Psi_{K,S}), \mathbf{0}_{3m \times 1}]'\right) / \sqrt{n} \\
= & h + O_p(\sqrt{\text{tr}(e_f)}) + O_p(\sqrt{K/n}) \\
= & h + o_p(1) \xrightarrow{d} N(0, \sigma^2 \bar{\mathbf{H}}),
\end{aligned}$$

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

Due to the increasing number of IVs $\sqrt{n}(\hat{\theta} - \theta)$ has the bias $\sqrt{n}b_{2SLS}$, when $K^2/n \rightarrow 0$ the bias term converges to zero and the sequence of IV matrices $\mathbf{Q}_{\mathbf{K}}$ gives the best IV estimator as $\sigma^2 \bar{\mathbf{H}}^{-1}$ reaches the efficiency lower bound for the IV estimators.

Under assumptions C1-C5, if $K^2/n \rightarrow 0$, then $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2 \bar{\mathbf{H}}^{-1})$.

In summary, having a sequence of IV matrices $\{Q_K\}$, condition $K/n \rightarrow 0$ is fundamental for the estimator to be consistent, while having $K^2/n \rightarrow 0$ provides the asymptotically best estimator because $\sigma^2 \bar{\mathbf{H}}^{-1}$ brings the lower bound for the IV estimators.

In this paper, we use 2SLS estimators and propose two innovations. First, we use two centralities, one for long-lived ties and one for short-lived ties in Q_2 (many-IVs 2SLS). Second, we take advantage of the longitudinal structure of our data and only include in the different instrumental matrices values lagged in time (i.e. observed in wave I). Let Q_{1L} and Q_{2L} denote the set of instruments Q_1 and Q_2 which only include variables in Wave I (i.e. lagged in time).

Note that $[G^L, G^S]$ has $2\bar{r}$ columns, so if we include Bonacich centralities for both short-lived and long-lived ties from each of the \bar{r} groups in $Q_{\mathbf{K}}$, then $2\bar{r}/K \rightarrow 0$. Hence, $K/n \rightarrow 0$ implies that $2\bar{r}/n = 2/\bar{s} \rightarrow 0$ where \bar{s} is the average group size. Then, as shown by Liu and Lee (2010) for the case of a single endogenous variable (i.e. coming from one interaction matrix), the average group size needs to be large enough, it should also be large relative to the number of groups because for the asymptotic efficiency it must be $K^2/n \rightarrow 0$ and it implies $(2\bar{r})^2/n = 2\bar{r}/\bar{s} \rightarrow 0$. If the network is not characterized by these properties, a bias correction should be used. Given the topology of the Add Health network, which is composed by quite a large number of relatively small networks, the best (feasible) estimator is the *bias-corrected* one

$$\hat{\theta}_{c2sls} = (\mathbf{Z}'P_{\mathbf{Q}_{\mathbf{K}}}\mathbf{Z})^{-1} \left[\mathbf{Z}'P_{\mathbf{Q}_{\mathbf{K}}}\mathbf{y} - \tilde{\sigma}^2 \left[\text{tr}(\tilde{\Psi}_{K,L}), \text{tr}(\tilde{\Psi}_{K,S}), \mathbf{0}_{3m \times 1} \right]' \right], \quad (15)$$

where $\tilde{\Psi}_{K,S} = \mathbf{P}_K \mathbf{G}^L \mathbf{M}(\tilde{\phi}^L, \tilde{\phi}^S)$ and $\tilde{\Psi}_{K,L} = \mathbf{P}_K \mathbf{G}^S \mathbf{M}(\tilde{\phi}^L, \tilde{\phi}^S)$ are estimated with initial \sqrt{n} -consistent estimators of $\tilde{\sigma}$, $\tilde{\phi}^L$ and $\tilde{\phi}^S$. This estimator adjusts the 2SLS estimator by the estimated leading order bias b_{2sls} , which is presented in Proposition 8.

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

Proof: We need to show that

$$\{\tilde{\sigma}^2 [\text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}}), \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}})]' - \sigma^2 [\text{tr}(\Psi_{K,L}), \text{tr}(\Psi_{K,S})]'\} / \sqrt{n} = o_p(1)$$

where $\tilde{\mathbf{M}} = \mathbf{M}(\tilde{\phi}^L, \tilde{\phi}^S)$. Given Proposition 8, this is quite straightforward since

$$\begin{aligned} & \{\tilde{\sigma}^2 [\text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}}), \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}})]' - \sigma^2 [\text{tr}(\Psi_{K,L}), \text{tr}(\Psi_{K,S})]'\} / \sqrt{n} \\ &= \sqrt{n}(\tilde{\sigma}^2 - \sigma^2) [\text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}}), \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}})]' / n \\ & \quad + \sqrt{n}\sigma^2 \left\{ \text{tr}[\mathbf{P}_K \mathbf{G}^L(\tilde{\mathbf{M}} - \mathbf{M})], \text{tr}[\mathbf{P}_K \mathbf{G}^S(\tilde{\mathbf{M}} - \mathbf{M})] \right\}' / n \\ &= \sqrt{n}(\tilde{\sigma}^2 - \sigma^2) [\text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}}), \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}})]' / n \\ & \quad + \sqrt{n}\sigma^2 [(\tilde{\phi}^L - \phi^L) \text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}} \mathbf{G}^L \mathbf{M}), 0]' / n \\ & \quad + \sqrt{n}\sigma^2 [(\tilde{\phi}^S - \phi^S) \text{tr}(\mathbf{P}_K \mathbf{G}^L \tilde{\mathbf{M}} \mathbf{G}^S \mathbf{M}), (\tilde{\phi}^L - \phi^L) \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}} \mathbf{G}^L \mathbf{M})]' / n \\ & \quad + \sqrt{n}\sigma^2 [0, (\tilde{\phi}^S - \phi^S) \text{tr}(\mathbf{P}_K \mathbf{G}^S \tilde{\mathbf{M}} \mathbf{G}^S \mathbf{M})]' / n \\ &= \mathbf{O}_p(\sqrt{K/n}) \\ &= o_p(1) \end{aligned}$$

because $\tilde{\mathbf{M}} - \mathbf{M} = \tilde{\mathbf{M}}[(\tilde{\phi}^L - \phi^L) \mathbf{G}^L \mathbf{M} + (\tilde{\phi}^S - \phi^S) \mathbf{G}^S \mathbf{M}]$, as a special case of Lemma C.11 in Lee and Liu (2010). ■

The 2SLS estimators of $\theta = (\phi^L, \phi^S, \beta)'$ considered in this paper are:

(i) *Finite-IV* : $\hat{\theta}_{2sls1} = (\mathbf{Z}' \mathbf{P}_1 \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{P}_1 \mathbf{y}$, where $\mathbf{P}_1 = \mathbf{Q}_1 (\mathbf{Q}'_1 \mathbf{Q}_1)_1^{-1} \mathbf{Q}'_1$ and \mathbf{Q}_1 contains the linearly independent columns of $\mathbf{J}[\mathbf{X}^*, \mathbf{G}\mathbf{X}^*, \mathbf{G}\mathbf{G}\mathbf{X}^*]$.

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

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

(iv) *Finite-IV lagged*: $\hat{\theta}_{2sls1L} = (\mathbf{Z}' \mathbf{P}_3 \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{P}_3 \mathbf{y}$, where $\mathbf{P}_3 = \mathbf{Q}_{1L} (\mathbf{Q}'_{1L} \mathbf{Q}_{1L})_{1L}^{-1} \mathbf{Q}'_{1L}$ and

\mathbf{Q}_{1L} contains the linearly independent and lagged in time columns of $\mathbf{J}[\mathbf{X}^*, \mathbf{GX}^*, \mathbf{GGX}^*]$.

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

(vi) *Bias-corrected lagged*:

$$\hat{\theta}_{c2slsL} = (\mathbf{Z}'\mathbf{P}_4\mathbf{Z})^{-1}\{\mathbf{Z}'\mathbf{P}_4\mathbf{y} - \tilde{\sigma}_{2sls1}^2[\text{tr}(\mathbf{P}_4\mathbf{G}^L\tilde{\mathbf{M}}), \text{tr}(\mathbf{P}_4\mathbf{G}^S\tilde{\mathbf{M}}), \mathbf{0}_{3m \times 1}]\}'\}$$

where $\tilde{\mathbf{M}} = (\mathbf{I} - \tilde{\phi}_{2sls1L}^L \mathbf{G}^L - \tilde{\phi}_{2sls1L}^S \mathbf{G}^S)^{-1}$, and $\tilde{\sigma}_{2sls1L}^2$, $\tilde{\phi}_{2sls1L}^L$ and $\tilde{\phi}_{2sls-1L}^S$ are \sqrt{n} -consistent initial estimators of σ^2 , ϕ^L and ϕ^S obtained by *Finite-IV lagged*.

Appendix C: Network Structure Indicators

Let N be a set of nodes with cardinality n . Let G be the adjacency matrix, whose generic element g_{ij} is equal to one if an edge (link) from j to i exists (here we consider indirect networks, so $g_{ij} = g_{ji}$). We consider the following network structure measures (Wasserman and Faust, 1994).

Density

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

Betweenness centrality

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

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

It assumes values in $[0, 1]$. At the network level:

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

where B_i^* is the maximum value of betweenness centrality among the nodes. This index is equal to one when a node has centrality equal to 1 and all others have zero centrality.

Closeness centrality

Let $d(i, j)$ be the shortest path between two nodes. For each node, closeness centrality is:

$$C_{2i} = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d(i, j)}.$$

At the network level:

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

Assortativity

Assortativity measures the correlation pattern in the degree distribution. If highly degree connected nodes are often linked with similar ones, it shows a positive sign. Let d_i be i 's number of links and $m = \sum_i \frac{d_i}{n}$ the average number of links among nodes. Assortativity is defined as:

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

Clustering coefficient

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

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

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

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

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

Appendix D: The DeGroot model

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

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

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

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

DeGroot (1974) assumes that agents update their beliefs by repeatedly taking *weighted averages* of their neighbors' beliefs with p_{ij} being the weight that agent i places on the current belief of agent j in forming his or her belief for the next period. To be more precise,

in the beginning, each agent $i \in N$ assigns a (relative) *weight* p_{ij} on each of his/her neighbors j such that $\sum_{j=1}^{j=n} p_{ij} = 1$. Observe that p_{ij} is the direct influence of agent j on i so that $p_{ij} = 0$ for $j \notin D(i)$.

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

$$p_{ij} = \frac{g_{ij}\pi_j}{\sum_{k=1}^{k=n} g_{ik}\pi_k} \quad (16)$$

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

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

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

or, in matrix form

$$\mathbf{b}^{(t)} = \mathbf{P} \mathbf{b}^{(t-1)} \quad (18)$$

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

$$\mathbf{b}^\infty = \lim_{t \rightarrow \infty} \mathbf{P}^t \mathbf{b}^{(0)} \quad (19)$$

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

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

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

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

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

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

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

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

where \mathbf{e} is the unique left eigenvector of matrix \mathbf{P} .

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

Assume that the network \mathbf{G} is undirected so that $g_{ij} = 1$ means that i and j pay attention to each other and define $p_{ij} = g_{ij}/d_i(\mathbf{G})$, where $d_i(\mathbf{G}) = \sum_{j \in N} g_{ij}$ is the degree of i . Then, under the same assumptions as in Proposition 8, the same convergence result holds where e_i is now given by:

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

This is a particular case of Proposition 8 where we assume that each agent equally split attention among his or her neighbors in an undirected network. In that case, Corollary

8 shows that social influence is proportional to the agent's degree. An interesting feature of Corollary 8 is that social influence is only determined by the *degree distribution* of the network and not by any structural properties of the network such as the average path length or the centrality of the network. However, Golub and Jackson (2010) shows that the *speed of convergence* is affected by a structural property of the network, namely *homophily*. They show that homophily does not alter agents' social influence and therefore has no effect on the long-term learning but significantly reduces the speed of convergence.

An example of convergence with the DeGroot model To illustrate the notations and the results in Proposition 8, let us consider the following network:

dtbpFU3.6729in3.6357in0ptFigure A1: Directed networkFigure

where the adjacency matrix is given by:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

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

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

This means, for example, that all agents put equal weight of all their neighbors. Assuming that the initial beliefs are given by: $\mathbf{b}^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$, it is easily verified that:

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

$$\mathbf{b}^{(2)} = \mathbf{P} \mathbf{b}^{(1)} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.278 \\ 0.417 \\ 0.25 \end{pmatrix}$$

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

$$\mathbf{b}^\infty = \lim_{t \rightarrow \infty} \mathbf{P}^t \mathbf{b}^{(0)} = \begin{pmatrix} 1/3 = 0.333 \\ 4/9 = 0.444 \\ 2/9 = 0.222 \end{pmatrix} \quad (20)$$

This means that

$$\mathbf{P}^t = \begin{pmatrix} 1/3 & 4/9 & 2/9 \\ 1/3 & 4/9 & 2/9 \\ 1/3 & 4/9 & 2/9 \end{pmatrix}$$

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

$$b_1^\infty = b_2^\infty = b_3^\infty = \frac{1}{3}b_1^{(0)} + \frac{4}{9}b_2^{(0)} + \frac{2}{9}b_3^{(0)} \quad (21)$$

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

$$b_1^\infty = b_2^\infty = b_3^\infty = \frac{1}{3} = 0.333$$

If there are two states of the world and if the consensus is on the state {It is worth continuing to study}, then this means that all students in this network agree that with 33.3 percent that it is worth continuing to study.

Appendix E: Bayesian Estimation

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

$$P(\mathbf{G}_r^L | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) = \prod_{i \neq j} P(g_{ij,r,t-1}^L | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}),$$

$$P(\mathbf{G}_r^S | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) = \prod_{i \neq j} P(g_{ij,r,t}^S | x_{ij,r}, z_{i,r}, z_{j,r}).$$

Following HL11 our prior distributions are

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

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

econometrician. It follows that the marginal posteriors are

$$\begin{aligned}
P(\mathbf{Z}_r | \mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L, \rho) &\propto \prod_{r=1}^{\bar{r}} \prod_i^{n_r} \phi(z_{i,r}) P(\mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L | \mathbf{Z}_r, \rho) \\
P(\omega | \mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L) &\propto \phi^{2K+3}(\omega, \omega_0, \Omega_0) \prod_{r=1}^{\bar{r}} P(\mathbf{G}_r^S, \mathbf{G}_r^L | \mathbf{Z}_r, \omega) \\
P(\phi^S, \phi^L | \mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L, \mathbf{Z}_r, \beta, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\propto \prod_{r=1}^{\bar{r}} P(\mathbf{Y}_r | \mathbf{G}_r^S, \mathbf{G}_r^L, \mathbf{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \\
P(\beta^* | \mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L, \mathbf{Z}_r, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \phi^S, \phi^L) &\propto \phi^{3K+2}(\tilde{\beta}, \tilde{\mathbf{B}}) \\
P(\sigma_\varepsilon^2, \sigma_{\varepsilon z} | \mathbf{Y}_r, \mathbf{G}_r^S, \mathbf{G}_r^L, \mathbf{Z}_r, \phi^S, \phi^L) &\propto \phi_T^2((\sigma_\varepsilon^2, \sigma_{\varepsilon z}), \sigma_0, \Sigma_0) \prod_{r=1}^{\bar{r}} P(\mathbf{Y}_r | \mathbf{G}_r^L, \mathbf{G}_r^S, \mathbf{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) \\
P(\eta_r | \mathbf{Y}_r, \mathbf{G}_r^L, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) &\propto \phi(\eta_r, \tilde{\eta}_r, \tilde{M}_r) \\
P(\sigma_\eta | \mathbf{Y}_r, \mathbf{G}_r^L, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\propto \nu\gamma\left(\frac{\zeta_0 + \bar{r}}{2}, \frac{\zeta_0 + \sum_{r=1}^{\bar{r}} \eta_r^2}{2}\right)
\end{aligned}$$

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

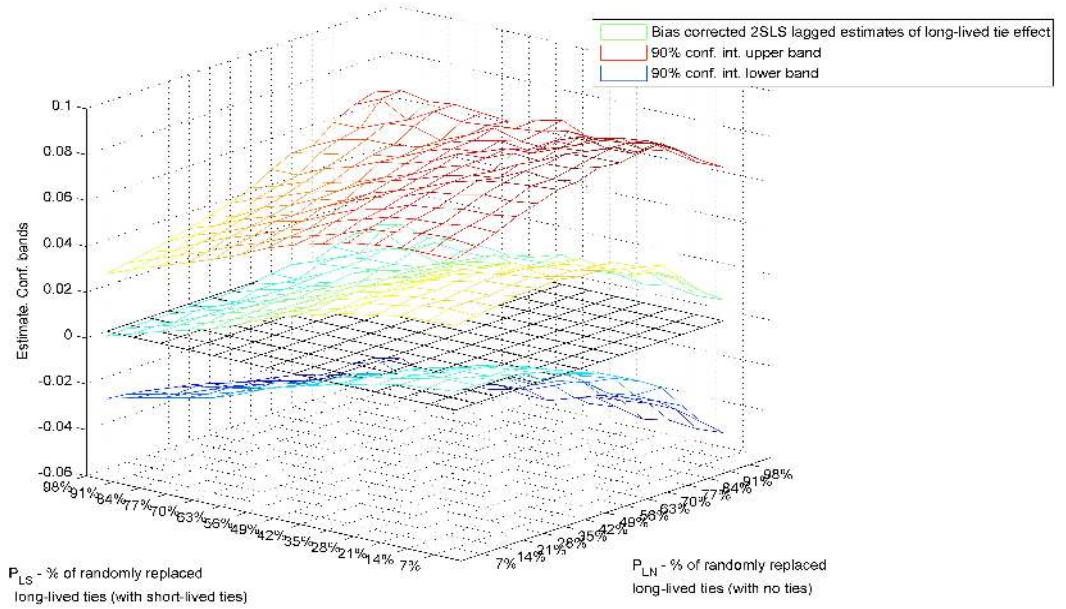
Sampling Algorithm We start our algorithm by picking $(\omega^{(1)}, \phi^{L(1)}, \phi^{S(1)}, \beta^{*(1)}, \sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}, \eta^{(1)})$ as starting values. For $\beta^{*(1)}, \eta^{(1)}, \phi^{L(1)}, \phi^{S(1)}$ we use OLS estimates, while we set the variances-covariances $\sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}$ at 0.³⁵ We ought to draw samples of $z_{i,r}^t$ from $P(z_{i,r}^t | \mathbf{Y}_r, \mathbf{G}_r^L, \mathbf{G}_r^S, \rho)$, $i = 1, \dots, n$. To do this, we first draw a candidate $\tilde{z}_{i,r}^t$ from a normal distribution with mean $z_{i,r}^{(t-1)}$, then we rely on a M-H decision rule: if $\tilde{z}_{i,r}^t$ is accepted we set $z_{i,r}^t = \tilde{z}_{i,r}^t$, otherwise $z_{i,r}^t = z_{i,r}^{(t-1)}$. Once all $z_{i,r}$ are sampled, we move to the sampling of β^* . By specifying a normal prior and a normal likelihood we can now easily sample β^t from a multivariate normal distribution. A diffuse prior for σ_ε^2 allows us to sample it from an inverse chi-squared distribution. We follow the Bayesian spatial econometric literature by sampling ϕ^S, ϕ^W from

³⁵The algorithm is robust to different starting values. However, speed of convergence may increase significantly.

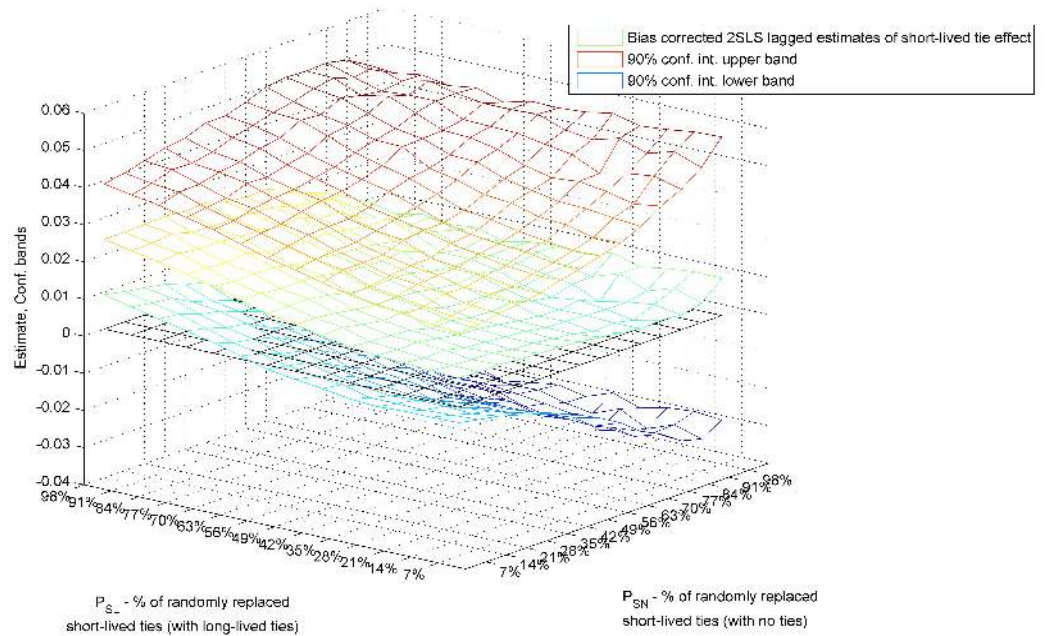
uniform distributions with support $[-\kappa_L, \kappa_L]$ and $[-\kappa_S, \kappa_S]$, as defined above. A M-H step is then performed over a normal likelihood: if accepted, then $\phi^{S^t} = \widetilde{\phi}^{S^t}$ and $\phi^{L^t} = \widetilde{\phi}^{L^t}$. For network fixed effects we deal again with normal prior and normal likelihood, so η is easily sampled from a multivariate normal. We sample $\sigma_\varepsilon^2, \sigma_{\varepsilon z}$ from a truncated bivariate normal over an admissible region Ξ such that the variance-covariance matrix is positive definite. Acceptance or rejection is determined by the usual M-H decision rule. At each of the M-H steps, the algorithm accepts the new values if the likelihood is higher than the current one.

Figure 1. Misspecification of long-lived ties and short-lived ties
Numerical simulation

Bias corrected 2SLS lagged estimates of long-lived tie effect (ϕ^L)



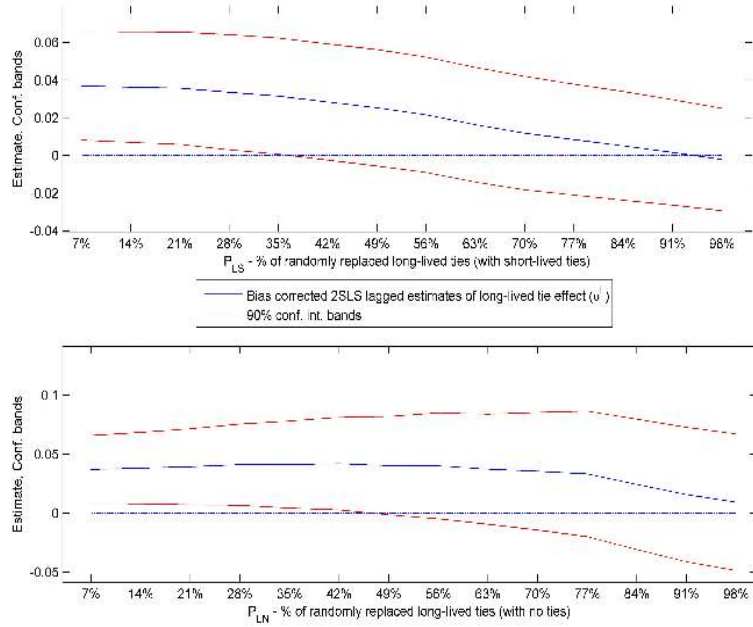
Bias corrected 2SLS lagged estimates of short-lived tie effect (ϕ^S)



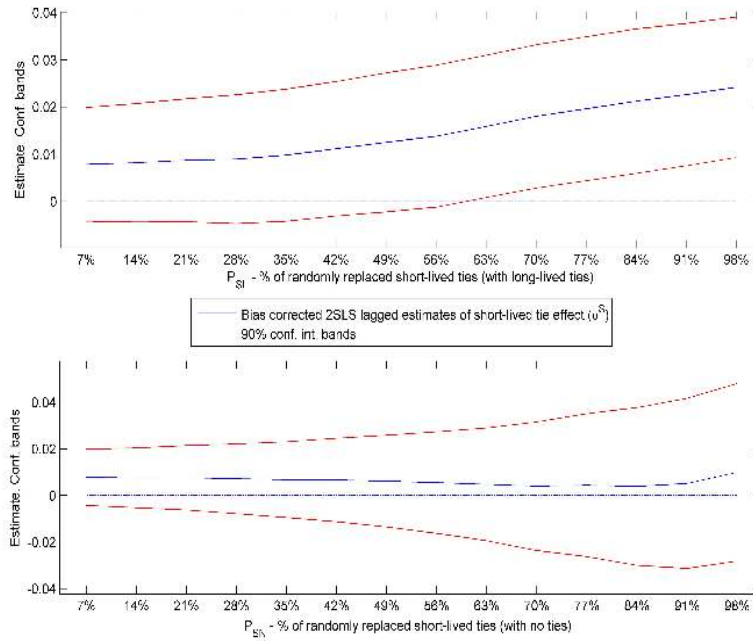
Notes: For each combination of replacement rates, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.

Figure 2. Simulation experiment
Single replacement effect

Bias corrected 2SLS lagged estimates of long-lived tie effect (ϕ^L)

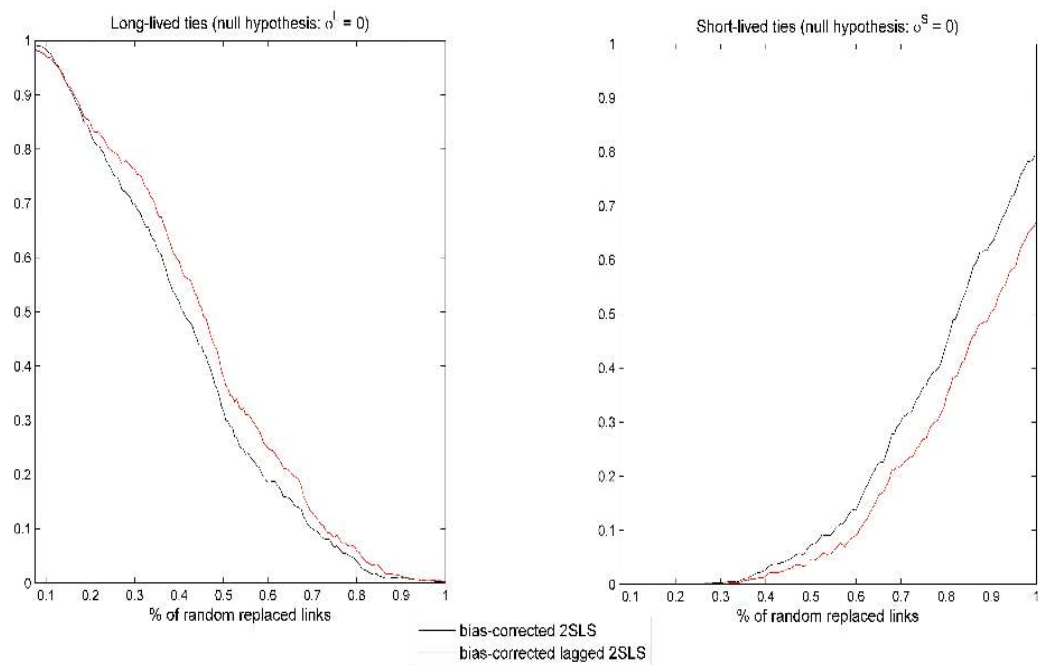


Bias corrected 2SLS lagged estimates of short-lived tie effect (ϕ^S)



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

Figure 3. Rejection rates of null hypothesis
Comparing estimators



Notes: the rejection rate of null hypothesis for long-lived and short-lived tie effects.

Table 1: Long-run peer effects

| Dep.Var. Years of Education | 2SLS | | | 2SLS Lagged | | |
|-----------------------------|-------------------------|-------------------------|-------------------------|------------------------|------------------------|------------------------|
| | Finite IV | Many IV | Bias Corrected | Finite IV | Many IV | Bias Corrected |
| Peer effects (ϕ) | 0.0057 *** (0.0020) | 0.0052 *** (0.0019) | 0.0052 *** (0.0019) | 0.0064 *** (0.0026) | 0.0058 *** (0.0020) | 0.0059 *** (0.0020) |
| Female | 0.9702 *** (0.2004) | 0.9718 *** (0.2004) | 0.9718 *** (0.2004) | 0.7383 (0.6123) | 1.0433 *** (0.2408) | 1.0434 *** (0.2408) |
| Black or African American | -0.1346 (0.4088) | -0.1341 (0.4088) | -0.1341 (0.4088) | -0.2933 (0.7722) | -0.1833 (0.4464) | -0.1830 (0.4464) |
| Other races | -0.3445 (0.2912) | -0.3447 (0.2911) | -0.3447 (0.2911) | -0.3873 (0.4561) | -0.2930 (0.3111) | -0.2927 (0.3111) |
| Religion Practice | 0.2515 *** (0.0540) | 0.2521 *** (0.0540) | 0.2521 *** (0.0540) | 0.2515 *** (0.1373) | 0.2521 *** (0.0599) | 0.2521 *** (0.0599) |
| Household Size | 0.0325 (0.0574) | 0.0327 (0.0574) | 0.0327 (0.0574) | 0.0247 (0.0900) | 0.0355 (0.0605) | 0.0355 (0.0605) |
| Parent education | 0.2890 *** (0.0969) | 0.2895 *** (0.0969) | 0.2895 *** (0.0969) | 0.2275 * (0.1528) | 0.2262 *** (0.1051) | 0.2262 *** (0.1051) |
| Mathematics score A | 1.3550 *** (0.2551) | 1.3578 *** (0.2551) | 1.3578 *** (0.2551) | 0.8917 * (0.5065) | 1.2388 *** (0.2816) | 1.2385 *** (0.2816) |
| Mathematics score B | 0.9691 *** (0.2399) | 0.9683 *** (0.2399) | 0.9683 *** (0.2399) | 0.6889 * (0.3966) | 0.8941 *** (0.2532) | 0.8941 *** (0.2532) |
| Mathematics score C | 0.5254 *** (0.2583) | 0.5272 *** (0.2583) | 0.5272 *** (0.2583) | 0.3422 (0.4019) | 0.5110 ** (0.2709) | 0.5107 ** (0.2709) |
| Mathematics score missing | 0.6214 (0.4286) | 0.6210 (0.4286) | 0.6210 (0.4286) | 1.1487 (0.8355) | 0.6678 (0.4471) | 0.6677 * (0.4471) |
| Resid. building qual. | 0.1771 ** (0.0966) | 0.1798 ** (0.0966) | 0.1797 ** (0.0966) | 0.0950 ** (0.1669) | 0.1264 ** (0.1053) | 0.1262 ** (0.1053) |
| Student grade | 0.4781 *** (0.0851) | 0.4766 *** (0.0850) | 0.4766 *** (0.0850) | 0.4051 *** (0.1654) | 0.5042 *** (0.0914) | 0.5043 *** (0.0914) |
| Children | -0.4171 *** (0.1668) | -0.4157 *** (0.1668) | -0.4157 *** (0.1668) | -1.5107 (1.7631) | -1.2429 ** (0.6263) | -1.2438** (0.6263) |
| Religion Practice (Wave 4) | 0.1511 *** (0.0514) | 0.1514 *** (0.0514) | 0.1514 *** (0.0514) | 1.0470 (1.1264) | 0.3429 ** (0.1873) | 0.3427 ** (0.1873) |
| Married | -0.1168 (0.1655) | -0.1166 (0.1655) | -0.1166 (0.1655) | -0.8750 (2.113) | -0.6020 (0.6381) | -0.6024 (0.6381) |
| Parental occupation dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Contextual effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Network fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| First stage F statistic | 1210.64 | 700.35 | | 1240.51 | 708.83 | |
| OIR test p-value | 0.503 | 0.461 | | 0.554 | 0.503 | |
| Observations | 1819 | 1819 | 1819 | 1819 | 1819 | 1819 |
| Networks | 116 | 116 | 116 | 116 | 116 | 116 |

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: 2SLS Lagged First Stage Results

| Dependent Variable: GY | | | | | |
|---------------------------|------------------------|-----------------------|--------------------------------------------|------------------------|------------------------|
| | X | GX | G^2X | G^3X | G^4X |
| Variables: X | Own | Peers | Peers of peers (Exclusion Restrictions) | | |
| Female | 0,4516 (0,4456) | 0,7482** (0,3483) | -0,2222 (0,1738) | -0,0209 (0,0422) | 0,0084 (0,0110) |
| Black or African American | -1,1913* (0,6487) | -0,3352 (0,4546) | 0,1265 (0,1264) | 0,0248 (0,0570) | -0,0061 (0,0071) |
| Other races | 0,1140 (0,6163) | 0,7957* (0,4818) | 0,0356 (0,2197) | -0,1779*** (0,0611) | 0,0176 (0,0162) |
| Religion Practice | -0,1296 (0,1202) | -0,0112 (0,0891) | -0,0007 (0,0431) | -0,0065 (0,0101) | 0,0012 (0,0024) |
| Household Size | -0,2091 (0,1355) | 0,1924** (0,0948) | 0,0392 (0,0458) | -0,0117 (0,0124) | 0,0013 (0,0029) |
| Parent education | -0,3902* (0,1966) | 0,6201*** (0,1520) | 0,1962*** (0,0625) | 0,0218 (0,0200) | -0,0062 (0,0043) |
| Mathematics score A | -1,5474* (0,5663) | 2,7857*** (0,4119) | 1,2068*** (0,1749) | 0,0480 (0,0463) | -0,0344*** (0,0093) |
| Mathematics score B | -1,0044* (0,5545) | 2,4891*** (0,3843) | 0,9199*** (0,1756) | -0,0298 (0,0461) | -0,0219*** (0,0090) |
| Mathematics score C | -1,0947* (0,5913) | 1,8637*** (0,4270) | 0,9009*** (0,2005) | 0,0658 (0,0479) | -0,0282*** (0,0114) |
| Mathematics score missing | -0,0857 (0,9780) | 1,8655*** (0,7666) | 0,4239 (0,3482) | -0,1896* (0,0891) | 0,0076 (0,0198) |
| Resid. building qual. | -0,0623 (0,1958) | -0,2335 (0,1545) | -0,0974 (0,0659) | -0,0019 (0,0204) | 0,0058 (0,0042) |
| Student grade | -0,5640*** (0,1559) | 1,0504*** (0,0838) | -0,1511*** (0,0263) | 0,0241** (0,0103) | -0,0028* (0,0016) |
| Network Fixed Effects | | | Yes | | |
| Number of Observations | | | 1819 | | |
| Number of Networks | | | 116 | | |

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

Table 3: Long and short-lived ties: Long-run effects

| Dep.Var. Years of Education | 2SLS | 2SLS Lagged |
|-------------------------------|-----------------------|-----------------------|
| Long-lived ties (ϕ^L) | 0.0317*** (0.0137) | 0.0345*** (0.0155) |
| Short-lived ties (ϕ^S) | 0.0062 (0.0055) | 0.0080 (0.0063) |
| Individual socio-demographic | yes | yes |
| Family Background | yes | yes |
| Protective Factors | yes | yes |
| Residential neighborhood | yes | yes |
| Contextual Effects | yes | yes |
| Network Fixed Effects | yes | yes |
| Observations | 1819 | 1819 |
| Networks | 116 | 116 |

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

Table 4: Short-lived ties in different grades (10 - 12)

| Dep.Var. Years of Education | 2SLS | |
|-----------------------------------------------|-----------------------|----------------------|
| | 2SLS | 2SLS Lagged |
| Long-lived ties (ϕ^L) | 0.0419*** (0.0187) | 0.0485** (0.0212) |
| Short-lived in lower grades (ϕ^{S_1}) | 0.0015 (0.0237) | 0.0027 (0.0331) |
| Short-lived in higher grades (ϕ^{S_2}) | 0.0011 (0.0207) | 0.0049 (0.0257) |
| Individual socio-demographic | yes | yes |
| Family Background | yes | yes |
| Protective Factors | yes | yes |
| Residential neighborhood | yes | yes |
| Contextual Effects | yes | yes |
| Network Fixed Effects | yes | yes |
| Observations | 628 | 628 |
| Networks | 41 | 41 |

Notes: We report bias-corrected 2SLS estimates.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Peer characteristics

| | Wave I | | | | Wave II | | | | Wave I and Wave II | | | |
|------------------------------|--------|------|-------|-------|---------|------|-------|-------|--------------------|------|-------|-------|
| | mean | std | min | max | mean | std | min | max | mean | std | min | max |
| Years of education | 14.96 | 3.48 | 9.00 | 23.00 | 14.62 | 3.35 | 9.00 | 23.00 | 14.91 | 3.39 | 9.00 | 23.00 |
| GPA | 2.89 | 0.87 | 0.00 | 4.85 | 2.91 | 0.94 | 0.00 | 4.81 | 2.88 | 0.97 | 0.00 | 4.82 |
| Female | 0.48 | 0.50 | 0.00 | 1.00 | 0.43 | 0.50 | 0.00 | 1.00 | 0.47 | 0.50 | 0.00 | 1.00 |
| Black or African American | 0.06 | 0.24 | 0.00 | 1.00 | 0.01 | 0.12 | 0.00 | 1.00 | 0.04 | 0.20 | 0.00 | 1.00 |
| Other races | 0.09 | 0.29 | 0.00 | 1.00 | 0.06 | 0.23 | 0.00 | 1.00 | 0.05 | 0.23 | 0.00 | 1.00 |
| Religion Practice | 3.98 | 1.82 | 1.00 | 7.00 | 3.88 | 1.89 | 1.00 | 7.00 | 3.84 | 1.86 | 1.00 | 7.00 |
| Household Size | 3.34 | 1.37 | 1.00 | 10.00 | 3.32 | 1.34 | 1.00 | 10.00 | 3.47 | 1.38 | 1.00 | 10.00 |
| Two married parents family | 0.75 | 0.44 | 0.00 | 1.00 | 0.77 | 0.42 | 0.00 | 1.00 | 0.80 | 0.40 | 0.00 | 1.00 |
| Parent education | 3.33 | 0.90 | 1.00 | 5.00 | 3.22 | 0.83 | 1.00 | 5.00 | 3.19 | 0.83 | 1.00 | 5.00 |
| Mathematics score A | 0.24 | 0.42 | 0.00 | 1.00 | 0.21 | 0.41 | 0.00 | 1.00 | 0.25 | 0.43 | 0.00 | 1.00 |
| Mathematics score B | 0.28 | 0.45 | 0.00 | 1.00 | 0.32 | 0.47 | 0.00 | 1.00 | 0.31 | 0.46 | 0.00 | 1.00 |
| Mathematics score C | 0.25 | 0.43 | 0.00 | 1.00 | 0.24 | 0.43 | 0.00 | 1.00 | 0.24 | 0.43 | 0.00 | 1.00 |
| Mathematics score missing | 0.08 | 0.27 | 0.00 | 1.00 | 0.08 | 0.28 | 0.00 | 1.00 | 0.07 | 0.26 | 0.00 | 1.00 |
| Residential building quality | 1.56 | 0.78 | 1.00 | 4.00 | 1.59 | 0.79 | 1.00 | 4.00 | 1.52 | 0.79 | 1.00 | 4.00 |
| Student grade | 10.44 | 0.50 | 10.00 | 11.00 | 10.43 | 0.50 | 10.00 | 11.00 | 10.40 | 0.49 | 10.00 | 11.00 |
| Children | 0.48 | 0.50 | 0.00 | 1.00 | 0.48 | 0.50 | 0.00 | 1.00 | 0.45 | 0.50 | 0.00 | 1.00 |
| Religion Practice (Wave 4) | 1.44 | 1.62 | 0.00 | 5.00 | 1.32 | 1.56 | 0.00 | 5.00 | 1.46 | 1.66 | 0.00 | 5.00 |
| Married | 0.50 | 0.50 | 0.00 | 1.00 | 0.50 | 0.50 | 0.00 | 1.00 | 0.49 | 0.50 | 0.00 | 1.00 |

Notes: Differences between means are never statistical significant at conventional levels of significance

Table 6: Network formation in Wave I and Wave II
Model (8) OLS estimation results

| VARIABLE | γ Coefficient | Std. error |
|------------------------------------------|----------------------|------------|
| Female | 0.0025 | (0.019) |
| Black or African American | -0.0107 | (0.044) |
| Other races | -0.0265 | (0.036) |
| Student grade | -0.0226 | (0.016) |
| Religion Practice | 0.0032 | (0.008) |
| Mathematics score A | -0.0063 | (0.017) |
| Mathematics score B | 0.0178 | (0.018) |
| Mathematics score C | 0.0122 | (0.019) |
| Mathematics score missing | -0.0030 | (0.022) |
| Parent education | -0.0073 | (0.013) |
| Household Size | 0.0125 | (0.008) |
| Parent occupation professional/technical | 0.0103 | (0.024) |
| Parent occupation manual | 0.0028 | (0.023) |
| Parent occupation office or sales worker | 0.0162 | (0.027) |
| Parent occupation other | 0.0460** | (0.022) |
| Two married parents family | -0.0026 | (0.020) |
| Residential building quality | 0.0085 | (0.012) |
| Network fixed effects | yes | |
| Chow test p value | 0.6083 | |
| Observations | 6,932 | |

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Wave I and Wave II network structure

| | Wave I | Wave II |
|------------------------------|--------|---------|
| Network structure indicators | | |
| Density | 0.0010 | 0.0007 |
| Betweenness | 0.0040 | 0.0055 |
| Closeness | 0.0131 | 0.0090 |
| Assortativity | 2.4105 | 2.9041 |
| Clustering coefficient | 0.0784 | 0.1516 |

Notes: Network structure indicators are described in Appendix D.

Table 8: Bayesian estimation

| | (a) homogeneous peer effects | | (b) heterogeneous peer effects | |
|-------------------------------|--------------------------------------------------|------------------------------------------|--------------------------------------------------|------------------------------------------|
| | 2SLS outcome eq. without link form. (1) | Bayesian estimation link form. (2) | 2SLS outcome eq. without link form. (4) | Bayesian estimation link form. (5) |
| Peer effects (ϕ) | 0.0045 *** (0.0017) | 0.0040** (0.0016) | 0.0097** (0.0042) | 0.0087** (0.0044) |
| Long-lived ties (ϕ^L) | | | | 0.0017 (0.0021) |
| Short-lived ties (ϕ^S) | | | | -0.1524 (0.0933) |
| Unobservables | | | | |
| $\sigma_{\epsilon z}$ | | 0.2303 (0.1887) | | |
| θ | | -3.0833*** (1.4227) | | |
| θ_L | | | | 0.7942*** (0.1262) |
| θ_S | | | | 0.5214*** (0.1168) |
| Individual socio-demographic | yes | yes | yes | yes |
| Family background | yes | yes | yes | yes |
| Protective factors | yes | yes | yes | yes |
| Residential neighborhood | yes | yes | yes | yes |
| Contextual effects | yes | yes | yes | yes |
| Network fixed effects | yes | yes | yes | yes |
| Observations | 932 | 433846 | 932 | 433846 |
| Networks | 33 | 33 | 33 | 33 |

Notes: Columns (2),(3),(5), and (6) report the means and the standard deviations of the posterior distributions of the parameters. We let our chain run for 200,000 iterations, discarding the first 20,000 iterations. Ergodicity of the Markov Chain is achieved.

Table 9: Directed networks

| Dep.Var. Years of Education | 2SLS | 2SLS Lagged |
|-------------------------------|-----------------------|-----------------------|
| Long-lived ties (ϕ^L) | 0.0393*** (0.0160) | 0.0474*** (0.0183) |
| Short-lived ties (ϕ^S) | 0.0049 (0.0063) | 0.0052 (0.0070) |
| Individual socio-demographic | yes | yes |
| Family Background | yes | yes |
| Protective Factors | yes | yes |
| Residential neighborhood | yes | yes |
| Contextual Effects | yes | yes |
| Network Fixed Effects | yes | yes |
| Observations | 1819 | 1819 |
| Networks | 116 | 116 |

Notes: We report bias-corrected 2SLS estimates.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Long and short-lived ties: Short-run effects

| Dep.Var. GPA | 2SLS | Bayesian estimation |
|-----------------------------------------|-----------------------|-----------------------|
| Long-lived ties (ϕ^L) | 0.0238*** (0.0097) | 0.0253*** (0.0093) |
| Short-lived ties (ϕ^S) | 0.0079* (0.0046) | 0.0121** (0.0051) |
| Unobservables ($\sigma_{\epsilon z}$) | | 0.0053 (0.1036) |
| Individual socio-demographic | yes | yes |
| Family Background | yes | yes |
| Protective Factors | yes | yes |
| Residential neighborhood | yes | yes |
| Contextual Effects | yes | yes |
| Network Fixed Effects | yes | yes |
| Observations | 1819 | 932 |
| Networks | 116 | 33 |

Notes: We report bias-corrected 2SLS estimates.

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

RECENTLY PUBLISHED “TEMI” (*)

- N. 1023 – *Understanding policy rates at the zero lower bound: insights from a Bayesian shadow rate model*, by Marcello Pericoli and Marco Taboga (July 2015).
- N. 1024 – *Accessorizing. The effect of union contract renewals on consumption*, by Effrosyni Adamopoulou and Roberta Zizza (July 2015).
- N. 1025 – *Tail comovement in option-implied inflation expectations as an indicator of anchoring*, by Sara Cecchetti, Filippo Natoli and Laura Sigalotti (July 2015).
- N. 1026 – *Follow the value added: bilateral gross export accounting*, by Alessandro Borin and Michele Mancini (July 2015).
- N. 1027 – *On the conditional distribution of euro area inflation forecast*, by Fabio Busetti, Michele Caivano and Lisa Rodano (July 2015).
- N. 1028 – *The impact of CCPs’ margin policies on repo markets*, by Arianna Miglietta, Cristina Picillo and Mario Pietrunti (September 2015).
- N. 1029 – *European structural funds during the crisis: evidence from Southern Italy*, by Emanuele Ciani and Guido de Blasio (September 2015).
- N. 1030 – *Female employment and pre-kindergarten: on the unintended effects of an Italian reform*, by Francesca Carta and Lucia Rizzica (September 2015).
- N. 1031 – *The predictive content of business survey indicators: evidence from SIGE*, by Tatiana Cesaroni and Stefano Iezzi (September 2015).
- N. 1032 – *Sovereign debt exposure and the bank lending channel: impact on credit supply and the real economy*, by Margherita Bottero, Simone Lenzu and Filippo Mezzanotti (September 2015).
- N. 1033 – *Does trend inflation make a difference?*, by Michele Loberto and Chiara Perricone (September 2015).
- N. 1034 – *Procyclicality of credit rating systems: how to manage it*, by Tatiana Cesaroni (September 2015).
- N. 1035 – *The time varying effect of oil price shocks on euro-area exports*, by Marianna Riggi and Fabrizio Venditti (September 2015).
- N. 1036 – *Domestic and international macroeconomic effects of the Eurosystem expanded asset purchase programme*, by Pietro Cova, Patrizio Pagano and Massimiliano Pisani (September 2015).
- N. 1037 – *Deconstructing the gains from trade: selection of industries vs. reallocation of workers*, by Stefano Bolatto and Massimo Sbracia (November 2015).
- N. 1038 – *Young adults living with their parents and the influence of peers*, by Effrosyni Adamopoulou and Ezgi Kaya (November 2015).
- N. 1039 – *Shoe-leather costs in the euro area and the foreign demand for euro banknotes*, by Alessandro Calza and Andrea Zaghini (November 2015).
- N. 1040 – *The macroeconomic effects of low and falling inflation at the zero lower bound*, by Stefano Neri and Alessandro Notarpietro (November 2015).
- N. 1041 – *The use of fixed-term contracts and the (adverse) selection of public sector workers*, by Lucia Rizzica (November 2015).
- N. 1042 – *Multitask agents and incentives: the case of teaching and research for university professors*, by Marta De Philippis (November 2015).
- N. 1043 – *Exposure to media and corruption perceptions*, by Lucia Rizzica and Marco Tonello (November 2015).
- N. 1044 – *The supply side of household finance*, by Gabriele Foà, Leonardo Gambacorta, Luigi Guiso and Paolo Emilio Mistrulli (November 2015).

(*) Requests for copies should be sent to:

Banca d’Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2013

- A. MERCATANTI, *A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with imperfect compliance*, Australian and New Zealand Journal of Statistics, v. 55, 2, pp. 129-153, **TD No. 683 (August 2008)**.
- F. CINGANO and P. PINOTTI, *Politicians at work. The private returns and social costs of political connections*, Journal of the European Economic Association, v. 11, 2, pp. 433-465, **TD No. 709 (May 2009)**.
- F. Busetti and J. MARCUCCI, *Comparing forecast accuracy: a Monte Carlo investigation*, International Journal of Forecasting, v. 29, 1, pp. 13-27, **TD No. 723 (September 2009)**.
- D. DOTTORI, S. I-LING and F. ESTEVAN, *Reshaping the schooling system: The role of immigration*, Journal of Economic Theory, v. 148, 5, pp. 2124-2149, **TD No. 726 (October 2009)**.
- A. FINICELLI, P. PAGANO and M. SBRACIA, *Ricardian Selection*, Journal of International Economics, v. 89, 1, pp. 96-109, **TD No. 728 (October 2009)**.
- L. MONTEFORTE and G. MORETTI, *Real-time forecasts of inflation: the role of financial variables*, Journal of Forecasting, v. 32, 1, pp. 51-61, **TD No. 767 (July 2010)**.
- R. GIORDANO and P. TOMMASINO, *Public-sector efficiency and political culture*, FinanzArchiv, v. 69, 3, pp. 289-316, **TD No. 786 (January 2011)**.
- E. GAIOTTI, *Credit availability and investment: lessons from the "Great Recession"*, European Economic Review, v. 59, pp. 212-227, **TD No. 793 (February 2011)**.
- F. NUCCI and M. RIGGI, *Performance pay and changes in U.S. labor market dynamics*, Journal of Economic Dynamics and Control, v. 37, 12, pp. 2796-2813, **TD No. 800 (March 2011)**.
- G. CAPPELLETTI, G. GUAZZAROTTI and P. TOMMASINO, *What determines annuity demand at retirement?*, The Geneva Papers on Risk and Insurance – Issues and Practice, pp. 1-26, **TD No. 805 (April 2011)**.
- A. ACCETTURO e L. INFANTE, *Skills or Culture? An analysis of the decision to work by immigrant women in Italy*, IZA Journal of Migration, v. 2, 2, pp. 1-21, **TD No. 815 (July 2011)**.
- A. DE SOCIO, *Squeezing liquidity in a "lemons market" or asking liquidity "on tap"*, Journal of Banking and Finance, v. 27, 5, pp. 1340-1358, **TD No. 819 (September 2011)**.
- S. GOMES, P. JACQUINOT, M. MOHR and M. PISANI, *Structural reforms and macroeconomic performance in the euro area countries: a model-based assessment*, International Finance, v. 16, 1, pp. 23-44, **TD No. 830 (October 2011)**.
- G. BARONE and G. DE BLASIO, *Electoral rules and voter turnout*, International Review of Law and Economics, v. 36, 1, pp. 25-35, **TD No. 833 (November 2011)**.
- O. BLANCHARD and M. RIGGI, *Why are the 2000s so different from the 1970s? A structural interpretation of changes in the macroeconomic effects of oil prices*, Journal of the European Economic Association, v. 11, 5, pp. 1032-1052, **TD No. 835 (November 2011)**.
- R. CRISTADORO and D. MARCONI, *Household savings in China*, in G. Gomel, D. Marconi, I. Musu, B. Quintieri (eds), *The Chinese Economy: Recent Trends and Policy Issues*, Springer-Verlag, Berlin, **TD No. 838 (November 2011)**.
- A. ANZUINI, M. J. LOMBARDI and P. PAGANO, *The impact of monetary policy shocks on commodity prices*, International Journal of Central Banking, v. 9, 3, pp. 119-144, **TD No. 851 (February 2012)**.
- R. GAMBACORTA and M. IANNARIO, *Measuring job satisfaction with CUB models*, Labour, v. 27, 2, pp. 198-224, **TD No. 852 (February 2012)**.
- G. ASCARI and T. ROPELE, *Disinflation effects in a medium-scale new keynesian model: money supply rule versus interest rate rule*, European Economic Review, v. 61, pp. 77-100, **TD No. 867 (April 2012)**.
- E. BERETTA and S. DEL PRETE, *Banking consolidation and bank-firm credit relationships: the role of geographical features and relationship characteristics*, Review of Economics and Institutions, v. 4, 3, pp. 1-46, **TD No. 901 (February 2013)**.
- M. ANDINI, G. DE BLASIO, G. DURANTON and W. STRANGE, *Marshallian labor market pooling: evidence from Italy*, Regional Science and Urban Economics, v. 43, 6, pp.1008-1022, **TD No. 922 (July 2013)**.
- G. SBRANA and A. SILVESTRINI, *Forecasting aggregate demand: analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework*, International Journal of Production Economics, v. 146, 1, pp. 185-98, **TD No. 929 (September 2013)**.
- A. FILIPPIN, C. V. FIORIO and E. VIVIANO, *The effect of tax enforcement on tax morale*, European Journal of Political Economy, v. 32, pp. 320-331, **TD No. 937 (October 2013)**.

- G. M. TOMAT, *Revisiting poverty and welfare dominance*, *Economia pubblica*, v. 44, 2, 125-149, **TD No. 651 (December 2007)**.
- M. TABOGA, *The riskiness of corporate bonds*, *Journal of Money, Credit and Banking*, v.46, 4, pp. 693-713, **TD No. 730 (October 2009)**.
- G. MICUCCI and P. ROSSI, *Il ruolo delle tecnologie di prestito nella ristrutturazione dei debiti delle imprese in crisi*, in A. Zazzaro (a cura di), *Le banche e il credito alle imprese durante la crisi*, Bologna, Il Mulino, **TD No. 763 (June 2010)**.
- F. D'AMURI, *Gli effetti della legge 133/2008 sulle assenze per malattia nel settore pubblico*, *Rivista di politica economica*, v. 105, 1, pp. 301-321, **TD No. 787 (January 2011)**.
- R. BRONZINI and E. IACHINI, *Are incentives for R&D effective? Evidence from a regression discontinuity approach*, *American Economic Journal : Economic Policy*, v. 6, 4, pp. 100-134, **TD No. 791 (February 2011)**.
- P. ANGELINI, S. NERI and F. PANETTA, *The interaction between capital requirements and monetary policy*, *Journal of Money, Credit and Banking*, v. 46, 6, pp. 1073-1112, **TD No. 801 (March 2011)**.
- M. BRAGA, M. PACCAGNELLA and M. PELLIZZARI, *Evaluating students' evaluations of professors*, *Economics of Education Review*, v. 41, pp. 71-88, **TD No. 825 (October 2011)**.
- M. FRANCESE and R. MARZIA, *Is there Room for containing healthcare costs? An analysis of regional spending differentials in Italy*, *The European Journal of Health Economics*, v. 15, 2, pp. 117-132, **TD No. 828 (October 2011)**.
- L. GAMBACORTA and P. E. MISTRULLI, *Bank heterogeneity and interest rate setting: what lessons have we learned since Lehman Brothers?*, *Journal of Money, Credit and Banking*, v. 46, 4, pp. 753-778, **TD No. 829 (October 2011)**.
- M. PERICOLI, *Real term structure and inflation compensation in the euro area*, *International Journal of Central Banking*, v. 10, 1, pp. 1-42, **TD No. 841 (January 2012)**.
- E. GENNARI and G. MESSINA, *How sticky are local expenditures in Italy? Assessing the relevance of the flypaper effect through municipal data*, *International Tax and Public Finance*, v. 21, 2, pp. 324-344, **TD No. 844 (January 2012)**.
- V. DI GACINTO, M. GOMELLINI, G. MICUCCI and M. PAGNINI, *Mapping local productivity advantages in Italy: industrial districts, cities or both?*, *Journal of Economic Geography*, v. 14, pp. 365-394, **TD No. 850 (January 2012)**.
- A. ACCETTURO, F. MANARESI, S. MOCETTI and E. OLIVIERI, *Don't Stand so close to me: the urban impact of immigration*, *Regional Science and Urban Economics*, v. 45, pp. 45-56, **TD No. 866 (April 2012)**.
- M. PORQUEDDU and F. VENDITTI, *Do food commodity prices have asymmetric effects on euro area inflation*, *Studies in Nonlinear Dynamics and Econometrics*, v. 18, 4, pp. 419-443, **TD No. 878 (September 2012)**.
- S. FEDERICO, *Industry dynamics and competition from low-wage countries: evidence on Italy*, *Oxford Bulletin of Economics and Statistics*, v. 76, 3, pp. 389-410, **TD No. 879 (September 2012)**.
- F. D'AMURI and G. PERI, *Immigration, jobs and employment protection: evidence from Europe before and during the Great Recession*, *Journal of the European Economic Association*, v. 12, 2, pp. 432-464, **TD No. 886 (October 2012)**.
- M. TABOGA, *What is a prime bank? A euribor-OIS spread perspective*, *International Finance*, v. 17, 1, pp. 51-75, **TD No. 895 (January 2013)**.
- G. CANNONE and D. FANTINO, *Evaluating the efficacy of european regional funds for R&D*, *Rassegna italiana di valutazione*, v. 58, pp. 165-196, **TD No. 902 (February 2013)**.
- L. GAMBACORTA and F. M. SIGNORETTI, *Should monetary policy lean against the wind? An analysis based on a DSGE model with banking*, *Journal of Economic Dynamics and Control*, v. 43, pp. 146-74, **TD No. 921 (July 2013)**.
- M. BARIGOZZI, CONTI A.M. and M. LUCIANI, *Do euro area countries respond asymmetrically to the common monetary policy?*, *Oxford Bulletin of Economics and Statistics*, v. 76, 5, pp. 693-714, **TD No. 923 (July 2013)**.
- U. ALBERTAZZI and M. BOTTERO, *Foreign bank lending: evidence from the global financial crisis*, *Journal of International Economics*, v. 92, 1, pp. 22-35, **TD No. 926 (July 2013)**.

- R. DE BONIS and A. SILVESTRINI, *The Italian financial cycle: 1861-2011*, *Cliometrica*, v.8, 3, pp. 301-334, **TD No. 936 (October 2013)**.
- G. BARONE and S. MOCETTI, *Natural disasters, growth and institutions: a tale of two earthquakes*, *Journal of Urban Economics*, v. 84, pp. 52-66, **TD No. 949 (January 2014)**.
- D. PIANESELLI and A. ZAGHINI, *The cost of firms' debt financing and the global financial crisis*, *Finance Research Letters*, v. 11, 2, pp. 74-83, **TD No. 950 (February 2014)**.
- J. LI and G. ZINNA, *On bank credit risk: systemic or bank-specific? Evidence from the US and UK*, *Journal of Financial and Quantitative Analysis*, v. 49, 5/6, pp. 1403-1442, **TD No. 951 (February 2015)**.
- A. ZAGHINI, *Bank bonds: size, systemic relevance and the sovereign*, *International Finance*, v. 17, 2, pp. 161-183, **TD No. 966 (July 2014)**.
- G. SBRANA and A. SILVESTRINI, *Random switching exponential smoothing and inventory forecasting*, *International Journal of Production Economics*, v. 156, 1, pp. 283-294, **TD No. 971 (October 2014)**.
- M. SILVIA, *Does issuing equity help R&D activity? Evidence from unlisted Italian high-tech manufacturing firms*, *Economics of Innovation and New Technology*, v. 23, 8, pp. 825-854, **TD No. 978 (October 2014)**.

2015

- M. BUGAMELLI, S. FABIANI and E. SETTE, *The age of the dragon: the effect of imports from China on firm-level prices*, *Journal of Money, Credit and Banking*, v. 47, 6, pp. 1091-1118, **TD No. 737 (January 2010)**.
- R. BRONZINI, *The effects of extensive and intensive margins of FDI on domestic employment: microeconomic evidence from Italy*, *B.E. Journal of Economic Analysis & Policy*, v. 15, 4, pp. 2079-2109, **TD No. 769 (July 2010)**.
- G. BULLIGAN, M. MARCELLINO and F. VENDITTI, *Forecasting economic activity with targeted predictors*, *International Journal of Forecasting*, v. 31, 1, pp. 188-206, **TD No. 847 (February 2012)**.
- A. CIARLONE, *House price cycles in emerging economies*, *Studies in Economics and Finance*, v. 32, 1, **TD No. 863 (May 2012)**.
- D. FANTINO, A. MORI and D. SCALISE, *Collaboration between firms and universities in Italy: the role of a firm's proximity to top-rated departments*, *Rivista Italiana degli economisti*, v. 1, 2, pp. 219-251, **TD No. 884 (October 2012)**.
- D. DEPALO, R. GIORDANO and E. PAPAPETROU, *Public-private wage differentials in euro area countries: evidence from quantile decomposition analysis*, *Empirical Economics*, v. 49, 3, pp. 985-1115, **TD No. 907 (April 2013)**.
- G. BARONE and G. NARCISO, *Organized crime and business subsidies: Where does the money go?*, *Journal of Urban Economics*, v. 86, pp. 98-110, **TD No. 916 (June 2013)**.
- P. ALESSANDRI and B. NELSON, *Simple banking: profitability and the yield curve*, *Journal of Money, Credit and Banking*, v. 47, 1, pp. 143-175, **TD No. 945 (January 2014)**.
- M. TANELI and B. OHL, *Information acquisition and learning from prices over the business cycle*, *Journal of Economic Theory*, 158 B, pp. 585-633, **TD No. 946 (January 2014)**.
- R. AABERGE and A. BRANDOLINI, *Multidimensional poverty and inequality*, in A. B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*, Volume 2A, Amsterdam, Elsevier, **TD No. 976 (October 2014)**.
- V. CUCINIELLO and F. M. SIGNORETTI, *Large banks, loan rate markup and monetary policy*, *International Journal of Central Banking*, v. 11, 3, pp. 141-177, **TD No. 987 (November 2014)**.
- M. FRATZSCHER, D. RIMEC, L. SARNOB and G. ZINNA, *The scapegoat theory of exchange rates: the first tests*, *Journal of Monetary Economics*, v. 70, 1, pp. 1-21, **TD No. 991 (November 2014)**.
- A. NOTARPIETRO and S. SIVIERO, *Optimal monetary policy rules and house prices: the role of financial frictions*, *Journal of Money, Credit and Banking*, v. 47, S1, pp. 383-410, **TD No. 993 (November 2014)**.
- R. ANTONIETTI, R. BRONZINI and G. CAINELLI, *Inward greenfield FDI and innovation*, *Economia e Politica Industriale*, v. 42, 1, pp. 93-116, **TD No. 1006 (March 2015)**.
- T. CESARONI, *Procyclicality of credit rating systems: how to manage it*, *Journal of Economics and Business*, v. 82, pp. 62-83, **TD No. 1034 (October 2015)**.
- M. RIGGI and F. VENDITTI, *The time varying effect of oil price shocks on euro-area exports*, *Journal of Economic Dynamics and Control*, v. 59, pp. 75-94, **TD No. 1035 (October 2015)**.

FORTHCOMING

- G. DE BLASIO, D. FANTINO and G. PELLEGRINI, *Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds*, *Industrial and Corporate Change*, **TD No. 792 (February 2011)**.
- A. DI CESARE, A. P. STORK and C. DE VRIES, *Risk measures for autocorrelated hedge fund returns*, *Journal of Financial Econometrics*, **TD No. 831 (October 2011)**.
- E. BONACCORSI DI PATTI and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register*, *Journal of Financial Intermediation*, **TD No. 848 (February 2012)**.
- M. MARCELLINO, M. PORQUEDDU and F. VENDITTI, *Short-Term GDP Forecasting with a mixed frequency dynamic factor model with stochastic volatility*, *Journal of Business & Economic Statistics*, **TD No. 896 (January 2013)**.
- M. ANDINI and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma*, *Journal of Economic Geography*, **TD No. 915 (June 2013)**.
- F. BRIPI, *The role of regulation on entry: evidence from the Italian provinces*, *World Bank Economic Review*, **TD No. 932 (September 2013)**.
- G. ALBANESE, G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values*, *Journal of Population Economics*, **TD No. 955 (March 2014)**.
- A. L. MANCINI, C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading*, *Review of Economics of the Household*, **TD No. 958 (April 2014)**.
- R. BRONZINI and P. PISELLI, *The impact of R&D subsidies on firm innovation*, *Research Policy*, **TD No. 960 (April 2014)**.
- L. BURLON, *Public expenditure distribution, voting, and growth*, *Journal of Public Economic Theory*, **TD No. 961 (April 2014)**.
- L. BURLON and M. VILALTA-BUFI, *A new look at technical progress and early retirement*, *IZA Journal of Labor Policy*, **TD No. 963 (June 2014)**.
- A. BRANDOLINI and E. VIVIANO, *Behind and beyond the (headcount) employment rate*, *Journal of the Royal Statistical Society: Series A*, **TD No. 965 (July 2015)**.
- G. ZINNA, *Price pressures on UK real rates: an empirical investigation*, *Review of Finance*, **TD No. 968 (July 2014)**.
- A. CALZA and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes*, *International Journal of Central Banking*, **TD No. 1039 (December 2015)**.