Estimating the effect of peers on non-cognitive skills of adolescents using friendship networks

Mridul Joshi*

December 1, 2021

Please find the statistical code used in the paper here

Abstract

There is now a significant body of literature in economics that emphasises the role of non-cognitive skills in later life outcomes. However, to the best of my knowledge, there is no published study till date that looks at the role of peers in the formation of non-cognitive skills. Using data on adolescent students and their peers from a nationally representative sample of schools in the United States, this paper investigates whether the non-cognitive outcomes of adolescents are influenced by the non-cognitive outcomes of their peers in school. Specifically, I look at emotional stability, extraversion and conscientiousness. I employ a spatial autoregressive model to separately identify the endogenous and contextual peer effects. I find evidence for presence of substantial and statistically significant endogenous peer effects in emotional stability and extraversion. I do not find evidence of peer effects in conscientiousness. My results are robust to measurement error as well as endogeneity in network formation.

Keywords: Peer effect, non-cognitive skills, friendship network, spatial autoregressive
JEL Classification: C31, D85, D91, I21, J24, Z13

---

*Paris School of Economics. Email: joshi.mridul9@gmail.com. I profusely thank my supervisor Margherita Comola and referee Liam Wren-Lewis for their support. I am grateful to Liam Wren-Lewis and Jeremie Gignoux for organising the Development Economics Research Seminar where I received feedback through the lifecycle of this dissertation. I also thank Aditi Bhowmick, Simoni Jain and Manpreet Singh for reading a preliminary version of this paper. The usual disclaimers apply. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health Web site (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Contents

List of Tables

1 Introduction

The idea that non-cognitive skills play a role in the labour market was introduced more than four decades ago by Bowles and Gintis (1976) who emphasized the importance of *employer-valued attributes* such as perseverance and punctuality. However, this strand of literature only gained importance three decades later following a series of influential papers by James Heckman and his collaborators. There is now a growing body of literature in economics that emphasizes the role of non-cognitive skills in the labour market. However, to the best of my knowledge, there is no published study till date that looks at the role of peers in the formation of non-cognitive skills.

In this paper, I try to fill this gap by looking at the role of peers in the development of non-cognitive skills amongst adolescents. To be more precise, I use methods from the networks and spatial econometrics literature to identify and estimate peer effects. I look at three primary non-cognitive skills or personality traits - Emotional Stability, Extraversion and Conscientiousness. These are three of the Big Five[1] personality traits (the other two being Openness and Agreeableness). My data does not contain the items required to construct the indices for Openness and Agreeableness and hence they are excluded from this analysis. The psychology literature predominantly refers to these outcomes as *personality traits* while the economics literature usually refers to them as *non-cognitive skills* although the use of the term *personality traits* is not uncommon in economics. Research has shown that they are malleable throughout adolescence (Cunha et al., 2006). Additionally, these outcomes are valued by the labour market, which also motivates the use of the word *skills*. Hence, I prefer the term *non-cognitive skills* as *personality traits* give the impression that these outcomes are fixed. However, given that I extensively cite from the literature in psychology, I use these terms interchangeably.

[1] Big Five or OCEAN is a commonly-used taxonomy in the psychology literature. These are predictive of various life outcomes such as years of schooling, job stability and propensity for crime etc over and above their effect through cognitive development. See Borghans et al. (2008) and Almlund et al. (2011).

For this empirical investigation, I use the National Longitudinal Survey of Adolescent to Adult Health survey, commonly referred to as Add Health. The Add Health data contains information on peers in the form of friendship nominations and we can use this data to reconstruct the entire social network in the sample schools. In addition to this, the Add Health data contains an exhaustive list of socio-economic, pyschological and health related variables that allow me to construct the outcomes and the covariates. To construct the outcomes of non-cognitive skills, I make use of item response theory (IRT) that models the relationship between an individual's response to an item and the underlying trait being measured.

While there are not many studies that have investigated peer effects in non-congnitve skills, peer effects in school outcomes such as grades and years of education is popular strand of literature in economics. Despite this, the evidence is not necessarily conclusive. This maybe attributed to the fact that identifying peer effects is rife with econometric issues (Manski, 1993; Angrist, 2014). A significant proportion of studies on this topic suffer from issues such as the reflection problem or correlated effects. The reflection problem arises because the peers' observed outcome is the result of the peers' background (Sacerdote, 2011) making it difficult to separate the endogenous peer effect (how the non-cognitive skills of an adolescent are affected by that of her peers) and contextual peer effects (how the characteristics of an adolescent's peers affect their own non-cognitive skills). Correlated effects are the product of shared environments (for instance, a common teacher) or selection effects (for instance, students select friends of same age). This conflates the peer effects with observed correlation in behaviours due to correlated effects. In most applied settings, randomised experiments make way for a clean identification. However, in the case of peer effects, their effectiveness is limited. First, randomisation does not solve the reflection problem to allow separating endogenous and contextual peer effects (Hsieh and Van Kippersluis, 2018). Second, the treatment to manipulate peer groups may not be successful. A classic example is

Carrell et al. (2013) where the students randomly assigned to peer groups segregated into homogenous groups.

In this paper, I employ an augmented spatial autoregressive (SAR) model and exploit the geometry of the students' social network to overcome the reflection problem and identify the endogenous as well as the contextual peer effect (Bramoullé et al., 2009; Lee et al., 2010). Since the peer groups are now individual specific and not completely overlapping, the characteristics of peers of peers serve as natural instruments for the peers' outcome (Bramoullé et al., 2009).

I employ network fixed effects to control for the correlated effects. However, even after controlling for network fixed effects, there may be unobserved characteristics that drive both friendship formation and outcome determination. I perform an intuitive test for endogenous network formation proposed by Goldsmith-Pinkham and Imbens (2013) and implemented by Patacchini et al. (2016). After conditioning on network fixed effects and a large vector of individual-level cotrols, there is very little to no evidence of network endogeneity.

I employ two estimation strategies. First, I perform an IV/2SLS estimation using the peers of peers characteristics as my instruments (Bramoullé et al., 2019). In addition to this, I perform a maximum likelihood estimation using a spatial autoregressive error term (Drukker et al., 2013; Lee et al., 2010). The likelihood function accounts for simultaneity by breaking the correlation between the independent variables and the error terms. My results show large and positive endogenous peer effects for two of my three outcome (emotional stability and extraversion). The endogenous peer effect for conscientiousness is noisily estimated and is not statistically different from zero. The results are robust to measurement error as well as endogeneity in network formation.

The rest of the paper is laid out in the following manner. Section 2 reviews the literature on peer effects in non-cognitive skills. Section 3 describes the data used, the construction of the outcomes and presents descriptive statistics. Section 4 is devoted to the identification and estimation strategies employed in this paper and section 5 presents the results. This is followed by three robustness checks in section 6. Section 7 concludes the paper.

2 Related literature

In this section I review the existing literature on peer effects in non-cognitive skills and relate it to my research question.

2.1 Personality traits and non-cogntive skills

There is some consensus amongst psychologists on the general taxonomy of generic personality traits commonly referred to as the Big Five or OCEAN comprising openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (emotional stability). These items encapsulate a range of related personality attributes that collectively form an individual's personality[2] (John and Srivastava, 1999). There are various instruments that have been designed to measure these traits. The most popular one is the 240-item NEO personality inventory (Costa Jr and McCrae, 2008) and various other revised and shorter versions of it that have been tested for their validity. Table 1 presents the Big Five traits as defined by the American Psychology Association Dictionary (VandenBos, 2007) [3]

It maybe easier to grasp non-cognitive skills such as grit, enthusiasm and perserverence as predictive of life outcomes compared to the broader (and more convoluted) Big Five traits. However, I choose to focus of the Big Five traits instead of more specific and narrowly defined traits as most metrics that are used to capture such personality traits can be mapped

[2]see Almlund et al. (2011) for their application to economics.

[3]I reproduced this from Almlund et al. (2011).

Table 1: The Big Five personality traits

Trait	**Definition**
Openness to experience	The tendency to be open to new aesthetic, cultural, or intellectual experiences.
Conscientiousness	The tendency to be organized, responsible, and hard-working
Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.
Agreeableness	The tendency of individuals to act in a cooperative, un-selfish manner.
Emotional stability	Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes. Neuroticism is its opposite.

Taken from the American Pyschology Association Dictionary (2007).

to the Big Five (Goldberg, 1993; Costa Jr and McCrae, 2008). For instance, traits such as grit, perserverence, ambition and self-discipline are strongly related to conscientiousness while traits like self-confidence, adventurousness and enthusiasm are related to extraversion. Traits like anxiousness, hostility, self-consciousness and impulsiveness are related to neuroticism and the absence thereof to emotional stability (Cunha et al., 2006; John and Srivastava, 1999).

Conscientiousness is a robust predictor of academic as well as labour outcomes. Using a meta-analytic framework, Poropat (2009) shows that the association between grades and conscientiousness are as big as those between grades and cognitive ability. Nyhus and Pons (2005) and Hogan and Holland (2003) amongst others show that conscientiousness is strongly associated with job performance but has a lower predictive power compared to IQ. Heckman et al. (2006) posit that adolescent facets of emotional stability like self-control and self-esteem predict adult earnings as much as cognitive ability. There is also empirical evidence that extraversion predicts better outcomes in the labour market (Fletcher, 2013; Deming, 2017). Last, Störmer and Fahr (2013) show that a one standard deviation increment in emotional stability leads to a 12 percent decline in absenteeism for men. Similarly, one standard deviation increase in agreeableness is associated with nine percent fewer absent days for men and a standard deviation increase in openness to experience is associated with 13 percent more absenteeism amongst women [4]. Thus, the literature has well-established that non-cognitive skills predict life outcomes [5]

Previous research has documented that these traits are somewhat stable in adulthood (Cobb-Clark and Schurer, 2012) but malleable during adolescence where they are influenced by their immediate environment and experiences (Cunha et al., 2006). The literature in psychology supports that that personality is not static but changes through the course of life, with most of these changes occuring in adolescence and emerging adulthood (Roberts et al., 2006). According to Hartup and Stevens (1997), during adolescence friends replace parents as confidantes and also "serve as an experimental field for later partner relationships" thus allowing individuals to develop their personalities based on the multifaceted interactions with their peers. Also, any relationship needs time and stability to exert influences on personality traits (Asendorpf and Wilpers, 1998) and thus adolescent frienships, by virtue of sharing a perpetual common environnment, may have a large impact on personality.

2.2 Peer effects in non-cognitive skills

The peer effects literature today has managed to circumvent some of the econometric challenges discussed briefly in the introduction by employing fixed effects or using some source of exogenous variation. For example, Hoxby (2000) and Hoxby and Weingarth (2005) employ changes in the gender mix across cohorts within a given school to show that an increase in the fraction of girls in the cohort lead to an increase in mean test scores of the peers. Another literature on peer effects relies on the random assignment of roommates or dormmates to identify the peer effect by inducing random variation in peer groups (Sacerdote, 2001; Glaeser

[4] Although a promising extension, I do not consider heterogeneous treatment effects by sex in this study.
[5] A review of all the evidence is beyond the scope of this paper. For a primer, see Cunha et al. (2006).

et al., 2003). However, these papers usually do not aim to separately identify the endogenous and the contextual peer effect and focus on the reduced form effect of changes in peer groups.

A third strand of peer effects literature uses social networks and is built around the seminal paper by Bramoullé et al. (2009) that presents conditions that allow the separate identification of endogenous and contextual peer effects. They show that the background characteristics of the friends' friends can serve as an instruments for the friends' endogenous outcome. Some of the papers that employ this strategy are Calvó-Armengol et al. (2009), Lin (2010), Fortin and Yazbeck (2015), Lin (2015) and Patacchini et al. (2017). However, this strategy is restricted by the availability of social network data (such as friendship nominations), which is not collected in standard school surveys. Add Health is one exception and it is no surprise that most of these studies, including mine, use this dataset.

I employ this social network approach for two main reasons. First, I am interested in the endogenous peer effect as it has the potential for social multipliers. Even a small change in a student's outcome can impact the peer group that then reflects back to the student. Second, it is sensible to assume that not all students in a class have an impact on a given student's outcomes, especially in the case of non-cognitive skills, as not all classmates are one's peers.

There is large literature on the effect of peers on school outcomes but the literature on peer effects in social outcomes is virtually missing. While there is no paper that directly looks at peer effects in non-cognitive skills there are a few papers that look at related outcomes and can help benchmark my results. Lin (2015) uses a spatial maximum likelihood approach to estimate peer effects in risky behaviour amongst adolescents. She finds evidence of large peer effects in skipping school, cigarette smoking, alcohol drinking and physical fighting. In a different setting, Zárate (2019) conducts a large-scale randomised experiment in boarding schools in Peru and finds that sociable peers have a positive effect on a student's social skills.

He shows that students that were randomly assigned to dormitories with more sociable peers show a 0.067 standard deviations higher social skills index. He also shows that the effect of more sociable peers on extraversion is 0.067 standard deviations and is statistically significant. This provides evidence that peer effects may be at play in determining the level of extraversion in a student.

Lavy and Sand (2019) exploit the Tel Aviv school application process wherein sixth-grade students list up to eight friends with whom they wish to attend school. and show that various types of friendships have positive effects on measures of well-being. The presence of friends in the class reduces fear from school violence and bullying and increases social satisfaction indicating that peer effects may be present in social outcomes. Last, Murphy and Weinhardt (2018) also provide evidence that non-cognitive skills are formed in peer groups. They show that academic rank during primary school has long-run impact on test scores and confidence, all of which cannot be explained by ability. They propose that students that are surrounded by students who perform worse than them improve non-cognitive skills such as grit, resilience and perseverance due to increased confidence. They dub this as the *Big Fish Little Pond* effect.

In addition to this, Golsteyn et al. (2017) show that peer personality influences academic outcomes. One may interpret them as contextual peer effects. They show that one standard deviation increase in average peer persistence raises grades by 0.018 standard deviation. The result is statistically significant. However, they do not find statistically significant effects of peer self-confidence, peer anxiety and peer risk attitude on test scores. In a similar vien, Hsieh and Van Kippersluis (2018) employ a spatial autoregressive model (SAR) and show that emotionally unstable individuals are more vulnerable to peer pressure in the case of smoking. They do not find any effect of conscientiousness in peer pressure in smoking. While the latter study does not provide direct evidence of peer effects in non-cognitive skills, it indicates they do play a role in social interactions.

There is some evidence of peer effects in non-cognitive skills in the psychology literature. However, the evidence is far from conclusive due to methodological issues associated with estimating peer effects. According to Mund and Neyer (2020), peer (relationship) effects in personality development are no unicorns but wild horses—difficult to observe in the wild but definitely real. The literature in pyschology proposes an interplay between what are known as personality effects - the effect of personality traits on the quality of peer (and other) relationships - and relationship effects - the effect of social relationship aspects on the development of personality (Mund et al., 2018). We are interested in the latter. According to the group socialisation theory proposed by Harris (1995), peer effects on personality development comprises two processes - assimilation and differentiation. Assimilation refers to the adoption of group norms that in turn affect the development of personality, resulting in increased similarity between peers over the course of time. On the other hand, differentiation refers to differences in group status and social comparisons that may result in increased dissimilarity of group members over time (Finn et al., 2017).

However, there is limited empirical evidence on these 'relationship effects'. van Zalk et al. (2020) observe dyadic relationships and show that friends' extraversion predicted increases in individuals's extraversion. This finding is supported by Ilmarinen et al. (2019) that finds that the network centrality predicts the development of extraversion. In a lab experiment, Reniers et al. (2017) employ a computerised financial risk taking task wherein they observed an increase in risk-taking in the presence of peers compared to when done alone. In the case of emotional stability, Mund and Neyer (2014) observe decreased neuroticism or higher emotional stability associated with close friendships. The results for conscientiousness and emotional stability as not as clear cut. On the other hand, according to Borghuis et al. (2017), adolescents show individual differences in the degree and direction of personality trait changes, with respect to the Big Five personality traits. They find no evidence for personality trait convergence or for correlated change in personality in dyadic friendship and sibling relationships. This may suggest that adolescent friends and siblings display changes in personality independently from each other and that their shared experiences do not have necessarily inform their personality traits. I take these results with a grain of salt as these papers do not have a clear identification strategy and rely on structural assumptions that are not watertight.

In contrast to Borghuis et al. (2017) and in line with van Zalk et al. (2020) and Mund and Neyer (2014), my results show the peer effects exist in extraversion and emotional stability (neuroticism).

3 Data and descriptive statistics

This section describes the data used, the construction of the outcomes and presents descriptive statistics.

3.1 National Longitudinal Survey of Adolescent to Adult Health

The National Longitudinal Survey of Adolescent to Adult Health, popularly known as the Add Health survey, is a nationally representative longitudinal survey of American adolescents in grades 7-12. It is a very comprehensive survey that contains an exhaustive set of social, pyschological, economic and health-related variables along with detailed questions on the respondents' schools, friendships, romantic relationships, family and neighbourhood of the respondents. This makes this dataset very popular amongst researchers studying various aspects of social interactions ranging from fast food consumption (Fortin and Yazbeck, 2015) and interracial relationships (Merlino et al., 2019) to bedtime decisions (Liu et al., 2017) and juvenile deliquency (Patacchini and Zenou, 2012).

Wave I of the survey ran from September 1994 to April 1995, wherein each student in the randomly sampled 130 high schools was administered a short questionnaire. An in-home survey of nearly 20000 students was also conducted with randomly drawn respondents from the sample schools. This questionnaire was more extensive and covered many sensitive topics like parent-child relationship, substance use, sexual activity and violence. The respondents were also asked to nominate up to five female and five male friends. This cohort was followed over the subsequent three waves of survey. The study also employed the extensive in-home questionnaire to create a *saturation sample* of 16 selected schools comprising 3000 students. Every student in these 16 schools answered the detailed questionnaire.

I employ the Wave I saturated sample of the Add Health survey in this study for two reasons. First, the questions used to create the outcome variables are present in the extensive in-home survey and second, using the saturated sample gives me access to the complete friendship network in these 16 schools. This allows me to see the non-cognitive skills of each student compared to that of their friends. Since, this is a subset of the in-school survey, an exhaustive set of control variables can be easily mapped to this data.

3.2 Outcomes

Waves I, II and III of the Add Health do not contain specific modules to measure non-cognitive skills or personality traits. However, some of the questions in wave I can potentially be used to construct measures of non-cognitive skills. I follow Young and Beaujean (2011) who use the Add Health data to create indices of emotional stability, extraversion and conscientiousness. These are the three outcomes that I use in this study. While I use the same items constituting each outcome measure as Young and Beaujean (2011), I do not follow their methods. The authors use lexical and factor-analytic methods to create their outcomes whereas I employ item response theory (IRT).

Item response theory is a model to express the relationship between an individual's response to an item and the underlying trait, θ, being measured (one of the three outcomes in the case at hand). θ is usually presented on a z-score scale with an effective range of -4 to +4. For an IRT model to work, it must be that a single trait accounts for the covariation amongst the items and the item characteristic function reflect the true relationship between the observed responses and the latent variable (Gray-Little et al., 1997). The items used to create the outcomes are likert scales hence a graded response IRT is employed. If the outcome levels for all items are given by $k = \{1, 2, 3, ..., K\}$ (with 1 referring to least conscientious and K to most conscientious), the probability of observing outcome k or higher for item i and person j is given by:

$$\Pr\left(X_{ij} \geq k \mid \theta_j\right) = \frac{\exp\left\{a_i\left(\theta_j - b_{ik}\right)\right\}}{1 + \exp\left\{a_i\left(\theta_j - b_{ik}\right)\right\}} \quad \theta_j \sim N(0,1)$$

where a_i is the discrimination of item i, b_{ik} is the threshold for item i and response category k, and θ_j is the latent trait of person j. The probability of observing outcome k can be calculated by differencing (Samejima, 1969). k parameters are estimated for each item - one discrimination parameter and $k-1$ threshold parameters. The items used to fit the IRT models are 5-category likert scales that range from *strongly disagree* to *strongly agree*. Table 2 lists these items [6].

Tables 3, 4 and 5 contain the parameter estimates for each of three outcomes estimated using the graded response IRT. Parameter a indicates how well the item discriminates between individuals with different levels of θ. It ranges from 0 to around 3 with higher values implying the item differentiates better amongst students (Gray-Little et al., 1997). The estimates of a for all three outcomes are significant. They are mostly high in magnitude

[6] see Young and Beaujean (2011) for a comparison of these items with the standard IPIP NEO-PI-R personality scale.

Table 2: Items for outcome indices from the Add Health wave I survey

Emotional stability
You have a lot of good qualities (H1PF 30)
You have a lot to be proud of (H1PF 32)
You like yourself just the way you are (H1PF33)
You feel like you are doing everything just about right (H1PF34)
You feel socially accepted (H1PF35)
You feel wanted and loved (H1PF36)
Extraversion
I feel close to people at school (S62B)
I feel like I am a part of this school (S62E)
I feel socially accepted (S62O)
Conscientiousness
When you have a problem to solve, the first things you do is get as many facts (H1PF18)
When you have a problem to solve, you usually think of different approaches (H1PF19)
When making decisions, you use a systematic method for judging alternatives (H1PF20)
After carrying out a solution, you analyze what went right and what went wrong (H1PF21)

The code is brackets is the question identifier in the AddHeath survey
Some of the questions are shortened for ease of reading

(between 2 and 3) for emotional stability and extraversion but lower in magnitude for conscientiousness. Parameter b represents a point at which a person with trait level $_j = b_{ik}$ has a 50% chance of responding in category k or higher. The estimates then imply that our items are informative at medium and high levels of θ but less informative for very low levels of θ.

Table 3: Emotional stability: IRT estimates for the 6 items

Item (Add Health)	a	b_1	b_2	b_3	b_4
H1PF30	2.264 (.094)	-.453 (.033)	1.491 (.048)	2.848 (.113)	4.287 (.348)
H1PF32	3.105 (.142)	-.227 (.028)	1.306 (.039)	2.217 (.071)	3.386 (.193)
H1PF33	2.033 (.078)	-.544 (.035)	.909 (.037)	1.665 (.055)	3.059 (.126)
H1PF34	1.891 (.073)	-1.307 (.048)	.535 (.033)	1.651 (.056)	3.350 (.148)
H1PF35	2.089 (.083)	-.779 (.037)	1.260 (.043)	2.272 (.078)	3.271 (.150)
H1PF36	2.202 (.089)	-.310 (.032)	1.475 (.048)	2.574 (.093)	3.673 (.207)

Table 4: Extraversion: IRT estimates for the 3 items

Item (Add Health)	a	b_1	b_2	b_3	b_4
S62B	2.188 (.107)	-1.108 (.0447)	.283 (.031)	1.270 (.047)	2.074 (.072)
S62E	2.576 (.145)	-.876 (.038)	.215 (.029)	1.087 (.041)	1.792 (.061)
S62O	1.750 (.080)	-1.114 (.049)	.515 (.037)	1.781 (.068)	2.767 (.111)

Table 5: Conscientiousness: IRT estimates for the 4 items

Item (Add Health)	a	b_1	b_2	b_3	b_4
H1PF18	1.906 (.084)	-1.146 (.046)	.872 (.039)	1.845 (.065)	3.264 (.142)
H1PF19	2.505 (.123)	-.862 (.037)	1.063 (.038)	1.968 (.064)	2.970 (.122)
H1PF20	1.958 (.086)	-1.512 (.053)	.484 (.033)	1.719 (.060)	3.163 (.135)
H1PF21	1.699 (.076)	-1.328 (.053)	1.027 (.044)	2.076 (.078)	3.726 (.185)

Table 6 shows the estimates of Chronbach's α that gives the internal validity (reliability) of our outcome indices or how closely related are the items as a group. A reliability coefficient of .70 or higher is considered acceptable (Taber, 2018). The coefficient for all the three outcomes is higher than this benchmark and very close to that of the standard personality instruments[7]

Table 6: Internal validity of index

Index	Chronbach's α
Emotional stability	0.84
Extraversion	0.75
Conscientiousness	0.75

We use the predicted values of θ from the IRT estimation of emotional stability, extraversion and conscientiousness as our outcome measures.

3.3 Descriptive statistics

Table 7 reports the summary statistics for our outcomes, covariates and other relevant characteristics of the sample. While θ ranges from -4 to +4, in our sample we only observe the measures of emotional stability, extraversion and conscientiousness between -2 and 4 with a mean value of -0.01, -0.03 and -0.01 for the three traits respectively.

One can observe that roughly half of the sample comprises females and the mean age is 15.44 years. In terms of racial composition, the share of white, black and asian people is 50%, 14% and 18% respectively. The mean household comprises 4.54 individuals and 93% of our respondents live with at least their mother. 75% of the mothers in the sample are atleast high school graduates. On a range of 1-4, the average GPA is 2.80. 15% of the respondents are in grades 7-8, 40% in grades 9-10 and 45% in grades 11-12.

Only 2% of the sample reports any difficulty in using their arms or their legs and on a scale of 1-5, the average self-reported health is 3.87. 47% of our sample smokes or has

[7] Caruso (2000) conducted a meta-analysis of all the NEO instruments, which shows that the average α coefficient is 0.88 for emotional stability and 0.83 for extraversion and conscientiousness.

had alcohol outside parental supervision. 45% of our sample also report attending religious services very often and 19% report they have never attended a religious service.

4 Identification and estimation strategy

In this section I discuss the peer effects model, the construction of the adjacency matrix, $\mathbf{G}$, the estimation methods employed and the threats to identification and their resolution.

Network Model

The baseline model of social interactions was studied extensively by Manski (1993) and is commonly referred to as Manski's linear-in-means model (MLIM). It is given by:

$$y_{ir} = \lambda E\left(y_r|r\right) + \beta x_{ir} + \gamma E\left(x_r|r\right) + \epsilon_{ir} \tag{1}$$

where y_{ir} denotes the outcomes of individual i in group r. x_{ir} denotes their characteristics and ϵ_{ir} denote the error terms. $E\left(y_r|r\right)$ and $E\left(x_r|r\right)$ denote the mean outcome and mean characteristics in group r. In terms of the setting at hand, λ denotes the endogenous peer effect (how the non-cognitive skills of an adolescent are affected by that of her peers) and γ denotes the contextual peer effect (how the characteristics of an adolescent's peers, for instance mother's education, impact their non-cognitive skills). The reduced form of this model is given by the following equation:

$$y_{ir} = \beta x_{ir} + \left(\frac{\lambda\beta + \gamma}{1-\lambda}\right) E\left(x_r \mid r\right) + \epsilon_{ir} \tag{2}$$

This equation illustrates the reflection problem: the endogenous peer effect λ cannot be separated from the contextual peer effect γ preventing peer effect identification. From equation 2, there is perfect collinearity between the mean outcome of the group and its mean

Table 7: Summary statistics

	Mean	SD	Min	Max
Emotional Stability	-0.01	0.92	-2	4
Extraversion	-0.03	0.84	-2	2
Conscientiousness	-0.01	0.87	-2	3
Female	0.51	0.50	0	1
Age	15.44	1.50	12	19
White	0.50	0.50	0	1
Black	0.14	0.34	0	1
Asian	0.18	0.39	0	1
Difficulty using arms or legs	0.02	0.15	0	1
Self-reported health	3.87	0.90	1	5
Attend religious services	2.89	1.18	1	4
Regularly smokes or drinks	0.47	0.50	0	1
Lives with mother	0.93	0.26	0	1
Parental education	2.36	0.76	0	4
School performance (GPA)	2.80	0.75	1	4
Household size	4.54	1.15	1	6
Grades 7-8	0.15	0.36	0	1
Grades 9-10	0.40	0.49	0	1
Grades 11-12	0.45	0.50	0	1
N	2308			

characteristics. Conceptualising peers in a social network can help circumvent this problem because the reference group therein is individual specific[8]. The approach is described below.

Suppose there are N individuals in the sample partitioned into r networks and n_r is the number of individuals in the rth network. Let $\mathbf{G_r^*}$ denote the $n_r \times n_r$ adjacency matrix of the rth network. We row-normalise $\mathbf{G_r^*}$ to get $\mathbf{G_r}$. Its element g_{ij} now takes the value $\frac{1}{n_r}$ when i and j are friends and 0 otherwise. Using the row-normalised adjacency matrix $\mathbf{G_r}$, we can write our peer effects in the following way:

$$y_{i,r} = \lambda \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + X_{i,r}\beta + \left(\sum_{j=1}^{n_r} g_{ij,r} X_{j,r}\gamma \right) + \eta_r + \epsilon_{i,r} \tag{3}$$

where y_{ir} denotes the non-cognitive skills of student i in network r and $X_{i,r}$ denotes the vector of characteristics (like age or sex) of individual i in network r. η_r denotes the network-specific fixed effects that controls for the shared environment faced by all the students in network r.

This can be written in matrix form as follows:

$$Y_r = \lambda \mathbf{G_r} Y_r + X_r\beta + \mathbf{G_r} X_r \gamma + \eta_r l_{n_r} + \epsilon_r \tag{4}$$

where l_{n_r} is an $n_r \times 1$ vector of ones. We can stack the data for all networks r using a block-diagonal matrix[9] to get:

$$Y = \lambda \mathbf{G} Y + X\beta + \mathbf{G} X\gamma + L\eta + \epsilon \tag{5}$$

Analogous to the MLIM model, λ gives the endogenous peer effect, γ gives the contextual peer effect and η gives the correlated effect.

[8]unless the network is complete i.e. all individuals are connected to each other.

[9]see Calvó-Armengol et al. (2009) and Liu et al. (2017) for details as well as for a microfoundation of this peer effects model.

In order to understand the source of identification, briefly assume the absence of the correlated effects. Bramoullé et al. (2009) show that the parameters of equation 4 are identified if the matrices $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G^2}$ are linearly independent. This condition is satisfied if networks are partially overlapping i.e. there are two individuals who are not friends but have a common friend. Intuitively, this condition breaks the perfect collinearity between the mean outcome of the group and its mean characteristics and solves the reflection problem because the characteristics of peers are no longer the same for all individuals as different individuals have different peer groups. Of course, this condition fails if the network is complete or if all individuals are linked to one another and peer effects are not identified. Fortunately, this condition is satisfied in almost all real world networks including the Add Health data.

4.1 Constructing the adjacency matrix

The Add Health survey asked respondents to nominate upto five female and five male friends. We exploit these nominations to contruct the network of peers in the data. To avoid adding more structure the nominations, I consider directed peers as is present in Add Health. If student A nominates student B but B does not nominate A, this means that A considers B their friend even if the nomination is reciprocated. Each element g_{ij} of the adjacency matrix for directed network $\mathbf{G}$ takes the value 1 if there is link from i to j i.e. i has nominated j as a friend and 0 otherwise. Note that $\mathbf{G}$ will not be symmetric. Students with no friendship nominations are excluded from the analysis.

Bramoullé et al. (2009) show that in the presence of correlated effects, the parameters of equation 5 are identified if and only if $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G^2}$ and $\mathbf{G^3}$ are linearly independent. This is because in order to get rid of the network fixed effect one has to do a within-network tranformation i.e. we need to average equation 4 over all the students in network r and substract it from each individual's equation. This adds the additional restriction. While this condition is more stringent than the one in the absence of correlated effect, it is satisfied in the data. Following Fortin and Yazbeck (2015), I compute the Belsley, Kuh and Welsch (Belsley et al., 2005) condition index to check for the presence of collinearity between these matrices. Linear independence holds if the index value is below 30. The condition index value is 5.63 in my data implying that the reflection problem does not cause identification issues.

4.2 Estimation methods

As discussed in the previous sections, Bramoullé et al. (2009) show that peer effects are identified if networks are partially overlapping i.e. there are two individuals who are not friends but are linked by two friends. In this case, the peers of peers do not directly affect a student but only have an indirect effect through the peers. This gives rise to natural exclusion restrictions in the model where the mean exogenous characteristics of peers of peers can be used as instruments for the mean peers' non-cognitive skills.

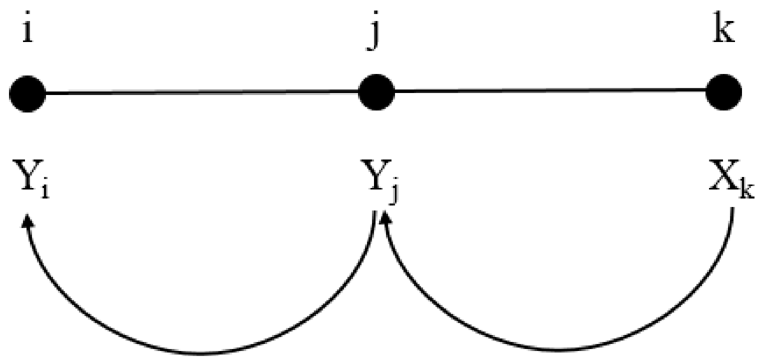
Figure 1 depicts students i and k who are only linked through student j but are otherwise not connected to each other (Patacchini et al., 2016). Again, for explanatory purposes assume away any correlated effects. Individual k affects individual i only through the common peer j. In other words, for student i, $\mathbf{G^2}X$[10] or x_k can instrument for $\mathbf{G}Y$ or y_j as x_k affects y_i only through its effect on y_j [11].

In addition to the IV estimation, I also estimate the model using a quasi-maximum

[10] See Appendix I for a primer on adjacency matrices.

[11] Lee (2003) establishes, under certain regularity conditions, the best generalized spatial two stage least squares estimator that is asymptotically efficient compared to other estimators in the class. He starts by estimating an IV regression with the peers of peers characteristics and then uses the resulting reduced form prediction to create an average of the predicted neighbor lags in the dependent variable. After creating this variable, he uses it as an instrument for the endogenous variable $\mathbf{G}Y$. However, my goal in this paper is to perform a 'credible' estimation using these 'naturally occurring' instruments without adding too much structure to the estimation. For this reason I forego the second step of Lee (2003) efficient estimation to focus on the relevance of peers of peers as instruments

Figure 1: Identification through intransitive triads



likelihood method (Lee et al., 2010) to estimate equation 5. However, the error term here is spatially autocorrelated:

$$\epsilon_r = \rho \mathbf{G}_r \epsilon_r + e_r \tag{6}$$

Lee et al. (2010) suggest that controlling ρ can reduce some of the bias from any omitted variables that may drive selection into groups. I briefly describe the maximum likelihood estimation here[12].

The following equation gives the reduced form of equation 5 where $\mathbf{S} = (\mathbf{I} - \lambda\mathbf{G})$, $\mathbf{Z} = (\mathbf{l}, \mathbf{X}, \mathbf{GX})$, $\delta_0 = (\eta, \beta', \gamma')$, and $\mathbf{R} = (\mathbf{I} - \rho\mathbf{G})$.

$$\mathbf{Y} = \mathbf{S}^{-1} \left(\mathbf{Z}\boldsymbol{\delta}_0 - \mathbf{R}^{-1}\boldsymbol{e}\right) \tag{7}$$

Given a row normalized weights matrix this implies that $\lambda \in (-1, 1)$. This is a standard assumption and identification fails in its absence (Bramoullé et al., 2009). A Cochrane-

[12]The exposition below borrows from Lee et al. (2010) and Norris (2018). For more details, refer to Lee et al. (2010)

Orcutt type transformation solves the problem of dependency between the observations.

$$\mathbf{RSY} = \mathbf{RZ}\delta_0 + \boldsymbol{e} \tag{8}$$

Assuming that $\epsilon \sim N\left(0, \sigma^2\right)$ the maximum log-likelihood estimator is

$$\mathcal{L} = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) + \sum_{s=1}^{\bar{s}} \ln\|\mathbf{S}\| + \sum_{s=1}^{\bar{s}} \ln\|\mathbf{R}\| - \frac{1}{2\sigma^2}\sum_{s=1}^{\bar{s}} \boldsymbol{e}'\boldsymbol{e} \tag{9}$$

Maximising with respect to $(\lambda, \beta', \gamma', \rho, \sigma)$ gives the maximum likelihood estimators.

Another way to intuitively grasp the identification of the endogenous peer effect is to rewrite the reduced form equation.

$$Y = S^{-1}\left(X\beta_{10} + GX\beta_{20}\right) + S^{-1}R^{-1}e \tag{10}$$

$$= X\beta_{10} + \sum_{j=0}^{\infty} \lambda^j G^{j+1} X\left(\beta_{20} + \lambda\beta_{10}\right) + S^{-1}R^{-1}e \tag{11}$$

This comes from the expansion of $\mathbf{S}^{-1}$, which yields

$$\mathbf{S}^{-1} = \sum_{j=0}^{\infty} (\lambda\mathbf{G})^j = \mathbf{I} + \lambda\mathbf{G} + \lambda^2\mathbf{G}^2 + \ldots \tag{12}$$

The effects of X on Y can be decomposed. The direct effect of X is captured by β_{10}, the effect due to peers is captured by $(\beta_{20} + \lambda\beta_{10})$, and that due to peers of peers is captured by $(\beta_{20} + \lambda\beta_{10})\,\lambda$, with the discount factor λ. Now that the peers can be distinguished from the peers of peers, the discount factor identifies the endogenous peer effect (Lee et al., 2010).

4.3 Threat to identification

Unobserved characteristics that predict the choice of peers and that are correlated with the outcomes (non-cognitive skills) may bias the estimates. To deal with this problem, we employ network fixed effects (Bramoullé et al., 2009). These fixed effects correct for the possibility of individuals with similar unobserved characteristics sorting into the same network. The idea is based on the assumption that such unobserved characteristics are common to all individuals in a given network (Patacchini et al., 2017).

While this is a reasonable solution, it may not take care of all forms of endogeneity. The adjacency matrix $\mathbf{G}$ may be endogenous even when controlling for the network fixed effects as failing to account for similarities in unobserved characteristics, similar behaviours may mistakenly be attributed to peer influence. Thus, I need to show that $\mathbf{G}$ is conditionally exogenous. In order to show that, I perform a test proposed by Goldsmith-Pinkham and Imbens (2013) that shows that if there is homophily in the unobserved characteristic, a testable implication of this problem of endogeneity would be to find a zero (implying no endogeneity) instead of a negative correlation between the predicted probability of forming a link (based on observable characteristics) and the unobserved similarity in pairs, as measured by the difference in residuals. I discuss this in detail in the next section.

Additionally, measurement error in the nomination data has the potential to bias the results. First, the Add Health survey only allowed students to nominate upto five male and five female students. If a sizeable proportion of the sample has more than five male or five female friends, it will induce measurement error in the adjacency matrix, $\mathbf{G}$. This is, however, not a problem in my sample as less than 10 percent of students nominate five male or female friends. There may be another type of measurement error where friendships are reciprocal but somehow missed in the nominations. To estimate the magnitude of this problem, I perform a robustness test where I consider all directed friendships to be reciprocal.

The results are virtually unchanged and only slightly larger in magnitude.

5 Results

In this section I perform a test for endogenous network formation proposed by Goldsmith-Pinkham and Imbens (2013) and discuss the estimation results.

5.1 Endogenous network formation

Goldsmith-Pinkham and Imbens (2013) propose an intuitive way to test any endogeneity in network formation. Under homophily, peers are similar in terms of observed characteristics as well as unobserved characteristics. Rewriting equation 4 with $\epsilon_r = \phi\nu_r + e_r$ where ν_r denotes a vector of unobserved individual characteristics and e_r denotes a vector of random disturbances. If $\nu \neq 0$, then $\mathbf{G_r}$ is endogenous.

$$Y_r = \lambda G_r Y_r + X_r\beta + G_r X_r\gamma + \eta_r l_{n_r} + \phi\nu_r + e_r \tag{13}$$

We can model the link formation process as follows:

$$g_{ij,r} = \alpha + \phi|x_i - x_j| - \mu|\nu_i - \nu_j| + \eta_r + u_{ij,r} \tag{14}$$

This is known as a latent space model (Hoff et al., 2002) and various studies of network formation have employed versions of this model (Fafchamps and Gubert, 2007; Udry and Conley, 2004; Santos and Barrett, 2010; Graham, 2014). Homophily in the observed characteristics imply that $\mu < 0$ i.e. the more similar the individuals are in their unobserved characteristics, the higher is the likelihood that they are peers.

This idea has testable implications. If there is homophily in the unobserved characteristic, a testable implication of this problem would be to find a zero (implying no endogeneity) instead of a negative correlation between the predicted probability of forming a link (based on observable characteristics), $\hat{g}_{ij,r}$ and the absolute value of the unobserved similarity in pairs, as measured by the difference in residuals, $|\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}|$ (Goldsmith-Pinkham and Imbens, 2013).

Table 8 presents the results from the estimation of equation 14. $g_{ij,r}$ takes the value 1 if i and j are peers and zero otherwise. The negative and statistically significant coefficients indicate the presence of homophily in observables. The number of possible links is 7.25 million which makes this regression highly powered to pick to pick up very small effects. We use the logit specification to predict the probability of observing a link between i and j denoted by $q_{ij,r}$.

I perform the endogeneity test as implemented by Patacchini et al. (2016). It has a simple premise. If we observe that for two students who are peers, the value of $q_{ij,r}$ is low, this means that the link formation between the two individuals is not explained by their observed characteristics. In this case the friendship is likely explained by the unobserved variables and, thus, one should observe low values of $|\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}|$ associated with low values of $q_{ij,r}$. Similarly, if we observe that for a dyad that is not linked, the value of $q_{ij,r}$ is high, the difference in their unobserved characteristics must be high in order to prevent them from making a link.

I regress $q_{ij,r}$ on $|\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}|$ in the data where a link is observed. I generate the quantiles of the distribution of $q_{ij,r}$ to determine high and low values of it. Table 9 presents the results for $g_{ij,r} = 1$. I consider two different thresholds for each of the three outcomes. Column 1 of table 9 defines low values of $q_{ij,r}$ as lower than the 30th percentile and column 2 defines them as lower than the 40th percentile. We see that there is no significant correlation for any of the outcomes[13].

[13] While a similar exercise can done for $g_{ij,r} = 0$, it may run into practical issues. One can assume that links are formed conditional on meeting the (prospective) peer. However, the total number of dyad combinations is large and it is likely that a significant proportion of these dyads may have similar characteristics but do not form a link for reasons not accounted by the model such as not meeting each other. Hence, I perform

Table 8: Dyadic regression

Dep Var: Probability of forming a link (g_{ij})

	Logit (1)	OLS (2)
main		
Sex	-0.249***	-0.002***
	[0.030]	[0.000]
Age	-0.476***	-0.003***
	[0.021]	[0.000]
White	-0.198***	-0.002***
	[0.050]	[0.000]
Black	-1.470***	-0.004***
	[0.074]	[0.000]
Asian	-1.519***	-0.004***
	[0.063]	[0.000]
School Performance	-0.265***	-0.002***
	[0.026]	[0.000]
Smokes or drinks	-0.392***	-0.002***
	[0.031]	[0.000]
Parental education	-0.073***	-0.001***
	[0.022]	[0.000]
Constant	-0.385**	0.266***
	[0.190]	[0.007]
N	725501	725501
School fixed effects	Yes	Yes
Parent occupation dummies	Yes	Yes
Extracurricular dummies	Yes	Yes

Notes: Standard errors (in parentheses).

The logit estimates are in terms of odd ratios.

* p<0.05, ** p<0.01, *** p<0.001

Table 9: Test of endogeneity

	Dep. var: $\lvert\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}\rvert$					
	Emotional Stability		Extraversion		Conscientiousness	
Threshold	$\tau = 0.30$ (1)	$\tau = 0.40$ (2)	$\tau = 0.30$ (3)	$\tau = 0.40$ (4)	$\tau = 0.30$ (5)	$\tau = 0.40$ (6)
Predicted probability	0.001	0.026	-0.085	-0.022	0.057	0.060
	[0.096]	[0.065]	[0.083]	[0.062]	[0.095]	[0.064]
Constant	1.605**	1.228***	0.931	0.692	1.312*	1.579***
	[0.682]	[0.455]	[0.591]	[0.440]	[0.674]	[0.446]
N	119	223	119	223	119	223
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors are in parentheses.
* p<0.05, ** p<0.01, *** p<0.001

I can also compute the probability that the predicted value of $q_{ij,r}$ lies above or below a threshold conditional on link formation i.e. $P(q_{ij} < \tau | g_{ij} = 1)$. I fail to predict link formation in approximately two percent of cases for $\tau = 40\%$ and approximately five percent of cases for $\tau = 30\%$. This exercise provides us with evidence of $\mathbf{G_r}$ being exogenous but it may not be conclusive. One must note that while the idea behind this test is powerful and intuitive, econometric analysis of the test and its statistical properties is absent from the peer effects literature Bramoullé et al. (2019). Column 2 table 8 shows that each observable component explains a very small part of network formation. There is no reason to believe that it is different for the unobserved characteristics. Thus, it may be the case that μ is close to zero and hence this test indicates little to no endogeneity in G_r. In order to account for a potentially non-zero ϕ, I specify the error term in my maximum likelihood specification as a spatially autocorrelated term. In addition to this, I employ a simple control function approach to model the endogenous network à la Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016) as a robustness check.

this analysis only for $g_{ij,r} = 1$.

5.2 Estimation results

I first estimate the model using IV/2SLS approach. However, while very intuitive it runs into some practical issues. The instruments based on the characteristics of peers of peers are possibly weak. The Donald-Cragg F statistic is lower than the rule-of-thumb value of 10 for all the three outcomes. The test statistic is below the critical value for the five percent worst case acceptable bias (21.36) and we can only reject the null of weak instrument for the 30 percent worst case acceptable bias (conscientiousness) and 20 percent worst case acceptable bias (emotional stability and extraversion). The issue is not unforeseen and has been discussed in Lee et al. (2010) and LeSage and Pace (2010) who focus on likelihood methods to circumvent the problem of weak instruments. However, this test may not be fullproof as there is no literature on the efficacy of this weak instrument test in a spatial set-up. On the other hand, the p-value from the Hansen J-statistic fails to reject the null for all the three models providing evidence that the overidentifying restrictions are met. The p-value from the Anderson LM statistic rejects the null for all the three models indicating that the instruments are not irrelevant. The tests are shown in table 10.

The possibility of weak instruments cannot be underplayed and this motivates the use of maximum likelihood estimation in this paper. This method, however, requires a normality assumption on the error term. I find that the magnitude and statistical significance of my peer effects estimates remains virtually the same irrespective of the estimation method. Although similar in magnitude and significance, I choose maximum likelihood as my preferred model due to the potential weak instruments problem with the IV/2SLS estimation. The estimation results are displayed in table 11.

Emotional stability

With the IV/2SLS approach, the estimation results show a large and positive endogenous effect of 0.288 [p-value < 0.001]. This means that the emotional stability of a student directly

Table 10: Instrument diagnostics

	Emotional Stability (1)	*Extraversion* (2)	*Conscientiousness* (3)
Anderson LM (p-value)	0.000	0.000	0.000
Donald-Cragg F statistic	5.293	7.451	4.622
Hansen J (p-value)	0.582	0.446	0.925

impacts the emotional stability of their peers. Additionally, the coefficient of proportion of female friends is -0.111 [p-value $<$ 0.05] implying a negative contextual peer effect in the sex of friends. The coefficient of peers' average parental education is small at 0.075 but statistically significant [p-value $<$ 0.05]. The results from the maximum likelihood estimation are very similar to that of IV/2SLS. The magnitude of the endogenous peer effect is 0.292 [p-value $<$ 0.001]. In addition to the contextual peer effect in sex of friends, one can observe a small contextual peer effect in the average parental education of peers. The magnitude is 0.074 [p-value $<$ 0.05], similar to that in the IV/2SLS estimation. ρ is negative and statistically significant. The negative spatial correlation among the errors may imply that the formation of friendships is not necessary driven by the common interest in the outcome at hand. Thus, the peer effect may be downward biased if this is not controlled for in the regressions (Lin, 2015).

Extraversion

The IV/2SLS estimates show a large and positive endogenous peer effect of 0.344 [p-value $<$0.001]. In addition to this, one can observe a positive and statistically significant contextual peer effect of magnitude 0.122 [p-value $<$0.01] in smoking or drinking. The maximum likelihood estimate of the endogenous peer effect is 0.353 [p-value $<$0.001], again very close to that obtained from IV/2SLS. As in the case of IV/2SLS, there is a contextual peer effect of magnitude 0.129 [p-value $<$0.01] in smoking and drinking. The ρ parameter is negative and statistically significant.

Conscientiousness

The endogenous peer effect from IV/2SLS estimation is 0.111. It is noisily estimated and statistically indistinguishable from zero. In addition to this, the results do not indicate the presence of any contextual peer effect as well. The endogenous peer effect from the ML estimation is 0.059 and statistically indistinguishable from zero. The magnitude is almost half of that from IV/2SLS estimation but the difference is not statistically significant. Like in the case of IV/2SLS, there is a striking absence of contextual peer effects.

The social multiplier

Endogenous peer effects have the potential to create social multiplier effects meaning that an increase in the outcome of some adolescents will lead to a stronger effects by influencing other adolescents. For instance, imagine the emotional stability of an adolescent exogenously increases. Due to the presence of endogenous peer effect, this will improve the emotional stability of the peers connected to the individual. The social multiplier is given by $\frac{1}{1-\lambda}$. In our data, the estimated social multiplier for emotional stability is 1.41 and that for extraversion is 1.55. These are large and may have interesting implications for education policy like ability tracking. On the other hand, the estimated social mutliplier for conscientiousness, with a magnitude of 1.06, is non-existent.

Table 11: Main results

	Emotional Stability		*Extraversion*		*Conscientiousness*	
	2SLS (1)	ML (2)	2SLS (3)	ML (4)	2SLS (5)	ML (6)
Proportion female friends	-0.111*	-0.093	0.039	0.043	-0.008	-0.007
	[0.062]	[0.061]	[0.056]	[0.054]	[0.060]	[0.060]
Peers' average age	-0.053	-0.055	-0.031	-0.020	-0.010	-0.010
	[0.039]	[0.038]	[0.035]	[0.034]	[0.037]	[0.037]
Proportion white peers	-0.007	-0.017	-0.054	-0.064	0.147	0.154
	[0.108]	[0.105]	[0.097]	[0.094]	[0.104]	[0.105]
Peers' average GPA	-0.041	-0.039	0.016	0.019	-0.022	-0.021
	[0.043]	[0.042]	[0.038]	[0.037]	[0.041]	[0.041]
Peers' average residential building quality	0.062	0.078	0.052	0.057	0.040	0.047
	[0.066]	[0.065]	[0.059]	[0.058]	[0.064]	[0.064]
Peers' average household size	-0.009	-0.012	-0.001	-0.011	-0.003	-0.003
	[0.027]	[0.026]	[0.024]	[0.023]	[0.026]	[0.026]
Peers' average parental education	0.075*	0.074*	-0.059	-0.048	0.051	0.052
	[0.042]	[0.041]	[0.038]	[0.037]	[0.041]	[0.041]
Proportion peers live with mother	-0.062	-0.113	0.020	0.019	-0.066	-0.070
	[0.109]	[0.106]	[0.097]	[0.095]	[0.105]	[0.105]
Proportion peers smoke or drink	-0.035	-0.014	0.122**	0.129**	-0.018	-0.015
	[0.061]	[0.059]	[0.054]	[0.053]	[0.058]	[0.058]
Endogenous peer effect	0.288***	0.292***	0.344***	0.353***	0.111	0.059
	[0.074]	[0.062]	[0.065]	[0.048]	[0.078]	[0.100]
ρ		-0.217***		-0.213***		-0.018
		[0.070]		[0.057]		[0.105]
N	2308	2308	2308	2308	2308	2308
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes	Yes	Yes	Yes
Parent occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sports dummies	Yes	Yes	Yes	Yes	Yes	Yes
Extracurricular dummies	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors (in parentheses).
All models include full set of own characteristics for which the contextual effects are estimated.
See appendix III for coefficients of individual characteristics.
* p<0.05, ** p<0.01, *** p<0.001

6 Robustness Checks

6.1 Control function approach

The control function approach involves estimating a selection equation followed by the outcome equation. The first stage is given by equation 14. Thus, the network formation equation is the selection equation in this set-up. I estimate it as a logit model. While the approach here is different, it is similar in spirit to the endogeneity test discussed in the previous section.

The idea behind this test is that the adjacency matrix G will be exogenous if the outcome equation is corrected for homophily by including the residuals from the network formation equation. The outcome equation where ν is the estimated residuals from the selection equation is given by:

$$Y_r = \lambda G_r Y_r + X_r \beta + G_r X_r \gamma + \eta_r l_{n_r} + \nu + \epsilon_r \tag{15}$$

This idea has been used by extensively discussed in Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016). However, both these papers employ a bayesian estimation strategy that is computationally very intensive. They derive a joint likelihood function of network formation and outcome equation. The resulting likelihood function cannot be solved analytically using MLE as it is difficult to integrate out the unobservable component. Hence, the authors prefer a bayesian estimation wherein they specify a prior distribution for the parameters and use MCMC to obtain draws from the posterior distribtution given the data. In contrast to that, I employ a simpler and less parametric version of this strategy. A similar strategy is implemented in Boguslaw (2017) but they use an additional exclusion restriction. In my set-up, the dyad-specific regressors used in the network formation model are naturally excluded from the outcome equation. Liu et al. (2017) also use a similar strategy but plug residuals from the outcome equation into the network formation equation.

I sum up my residuals for each individual and plug them in the outcome equation. Table

12 shows the estimates from the outcome equation that includes the selection term. The endogenous peer effect estimates from emotional stability and extraversion are very close to those estimated without a selection term. The endogenous peer effect for emotional stability declines by 0.003 and the one for extraversion declines by 0.009. Both remain statistically significant at the 1 percent level. Like in the main result, the endogenous peer effect estimate for conscientiousness increases by 0.018 but it plagued by a lot of noise and hence not interpretable.

6.2 Undirected networks

In this paper I assumed that friendships are directed i.e. each element g_{ij} of the adjacency matrix for directed network $\mathbf{G}$ takes the value 1 if there is link from i to j i.e. i has nominated j as a friend and 0 otherwise. However, it may be the case that friendships are reciprocal but were missed during the nomination survey due to enumerator error or miscalculation on the part of the respondent. The goal of this exercise is to observe the sensitivity of my results to a possible measurement error in network formation through nominations. Now, I assume that all friendships are symmetric even if just one friend nominates the other i.e. $g_{ij} = g_{ji}$.

The results of this exercise are given in table 13. The results are only slightly larger in magnitude and remain statistically significant.

6.3 Randomly assigned peers

As another robustness check, I perform a falsification test proposed by Lin (2015) that uses randomly assigned peers. I reassign peers randomly but without changing the number of links overall to any bias due to network density. Table 14 presents the results.

None of the estimates of endogenous peer effect are signficantly different from zero. This provides some assurance that the results are not driven by confounding variables.

Table 12: Control function approach

	Emotional Stability (1)	*Extraversion* (2)	*Conscientiousness* (3)
Proportion female friends	-0.093	0.043	-0.007
	[0.061]	[0.054]	[0.060]
Peers' average age	-0.054	-0.018	-0.009
	[0.038]	[0.034]	[0.037]
Proportion white peers	-0.019	-0.068	0.155
	[0.105]	[0.094]	[0.105]
Peers' average GPA	-0.040	0.019	-0.021
	[0.041]	[0.037]	[0.041]
Peers' average residential building quality	0.073	0.050	0.045
	[0.065]	[0.058]	[0.064]
Peers' average household size	-0.012	-0.011	-0.003
	[0.026]	[0.023]	[0.026]
Peers' average parental education	0.077*	-0.045	0.053
	[0.041]	[0.037]	[0.041]
Proportion peers live with mother	-0.130	-0.002	-0.078
	[0.106]	[0.095]	[0.105]
Proportion peers smoke or drink	-0.013	0.130**	-0.015
	[0.059]	[0.053]	[0.058]
Endogenous peer effect	0.286***	0.340***	0.064
	[0.062]	[0.048]	[0.097]
ρ	-0.214***	-0.206***	-0.025
	[0.070]	[0.057]	[0.101]
N	2308	2308	2308
Individual characteristics	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes
Parent occupation dummies	Yes	Yes	Yes
Sports dummies	Yes	Yes	Yes
Extracurricular dummies	Yes	Yes	Yes

Notes: Standard errors (in parentheses).

All models include full set of own characteristics for which the contextual effects are estimated.

* p<0.05, ** p<0.01, *** p<0.001

Table 13: Peer effects with undirected networks

	Emotional Stability (1)	*Extraversion* (2)	*Conscientiousness* (3)
Proportion female friends	-0.100*	0.033	-0.008
	[0.060]	[0.053]	[0.059]
Peers' average age	-0.056	-0.010	-0.008
	[0.037]	[0.033]	[0.037]
Proportion white peers	-0.029	-0.072	0.158
	[0.103]	[0.092]	[0.103]
Peers' average GPA	-0.045	0.019	-0.021
	[0.041]	[0.036]	[0.041]
Peers' average residential building quality	0.084	0.049	0.056
	[0.064]	[0.057]	[0.064]
Peers' average household size	-0.011	-0.009	-0.004
	[0.025]	[0.023]	[0.025]
Peers' average parental education	0.071*	-0.055	0.053
	[0.041]	[0.036]	[0.041]
Proportion peers live with mother	-0.121	0.046	-0.075
	[0.105]	[0.093]	[0.104]
Proportion peers smoke or drink	-0.011	0.114**	-0.013
	[0.058]	[0.052]	[0.058]
Endogenous peer effect	0.303***	0.364***	0.122
	[0.063]	[0.049]	[0.126]
ρ	-0.246***	-0.245***	-0.098
	[0.071]	[0.058]	[0.133]
N	2308	2308	2308
Individual characteristics	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes
Parent occupation dummies	Yes	Yes	Yes
Sports dummies	Yes	Yes	Yes
Extracurricular dummies	Yes	Yes	Yes

Notes: Standard errors (in parentheses).
All models include full set of own characteristics for which the contextual effects are estimated.
* p<0.05, ** p<0.01, *** p<0.001

Table 14: Peer effects with randomly assigned peers

	Emotional Stability (1)	*Extraversion* (2)	*Conscientiousness* (3)
Endogenous peer effect	-0.021	-0.025	-0.036
	[0.040]	[0.040]	[0.040]
N	2308	2308	2308
Individual characteristics	Yes	Yes	Yes
Contextual peer effects	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes
Parent occupation dummies	Yes	Yes	Yes
Sports dummies	Yes	Yes	Yes
Extracurricular dummies	Yes	Yes	Yes

Notes: Standard errors (in parentheses).
All models include full set of own characteristics and contextual effects.
* p<0.05, ** p<0.01, *** p<0.001

7 Conclusion

Despite substantial research on the role of non-cognitive skills on life outcomes, their importance has always been understated. This may be attributed to the absence of objective and tangible metrics of such skills vis-à-vis cognitive skills that can be objectively, albeit imperfectly, measured using academic test scores. While there are instruments designed to capture non-cognitive skills as well, they are less understood due to the latent nature of these skills and there is little consensus amongst researchers on the best set of items to capture each skill. Thus, in comparison to cognitive skills, they have made little impact in the policy space. Policy debates concerning tracking and affirmative action in schools revolve around the presence of peer effects in learning. However, they fail to take non-cognitive peer effects into account, which are valued equally by the labour market.

In this paper, I present evidence for the presence of a social multiplier in emotional stability and extraversion. To the best of my knowledge this is the first paper that documents the presence of endogenous peer effects in non-cognitive skills. My results are robust to endogenous network formation as well as potential measurement error in the friendship network. The presence of an endogenous peer effect has important implications for education policy. For instance, policies like ability tracking and classroom seating can be designed taking into account the non-cognitive skills that are valued in the labour market. Like in the case of cognitive skills, the social multiplier here implies that the increase in the non-cognitive skills of student will influence the non-cognitive skills of their peers, which, in turn, will affect those of their peers, and so on.

There are, however, a few limitations to this study. First, while I was able to construct internally valid indices for emotional stability, extraversion and conscientiousness by using questions, the absence of a standard module on non-cognitive skills in the Add Health data leaves something to be desired. Second, missing data in friendship nominations maybe a concern as individuals with no nominations were dropped from the sample. The robustness check with reciprocal data, which allows me to regain many of the lost observations, somewhat assuages this concern as the results from analysing the data under the assumption of reciprocal friendships are similar to those assuming directed friendships.

There are many possible extensions to this paper that I hope to explore in the future. It is likely that peer effects are more prominent in friendships that a long-lived. Indeed, Patacchini et al. (2017) finds that peer effects in education are larger for long-lived friendships. I can exploit the multiple waves of the survey to identify long-lived friendships and estimate the magnitude of peer effects therein. Another natural extension is heterogeneous peer effects by gender and the rank of friends. There is some evidence that the labour market valuation of different non-cognitive outcomes varies by gender (Nyhus and Pons, 2005). Hence, it is likely that peer effects may play out differently by gender. In the case of rank of friends, it is intuitive that a student's closest friends will have the largest effect on their personality and non-cognitive skills. Last, it will also be interesting to investigate the persistence of these peer effects and their bearing on later life outcomes like income, job satisfaction and criminal behaviour.

References

Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.

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

Asendorpf, J. B. and Wilpers, S. (1998). Personality effects on social relationships. *Journal of personality and social psychology*, 74(6):1531.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons.

Boguslaw, J. (2017). The key player in disruptive behavior: Whom should we target to improve the classroom learning environ ment. *Opublicerat manuskript.*

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

Borghuis, J., Denissen, J. J., Oberski, D., Sijtsma, K., Meeus, W. H., Branje, S., Koot, H. M., and Bleidorn, W. (2017). Big five personality stability, change, and codevelopment across adolescence and early adulthood. *Journal of Personality and Social Psychology*, 113(4):641.

Bowles, S. and Gintis, H. (1976). Schooling in capitalist america. *New York: Basic.*

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

Bramoullé, Y., Djebbari, H., and Fortin, B. (2019). Peer effects in networks: A survey.

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

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Caruso, J. C. (2000). Reliability generalization of the neo personality scales. *Educational and Psychological Measurement*, 60(2):236–254.

Cobb-Clark, D. A. and Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1):11–15.

Costa Jr, P. T. and McCrae, R. R. (2008). *The Revised NEO Personality Inventory (NEO-PI-R)*. Sage Publications, Inc.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Drukker, D. M., Prucha, I. R., and Raciborski, R. (2013). Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *The Stata Journal*, 13(2):221–241.

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation. *American Economic Review*, 97(2):75–79.

Finn, C., Zimmermann, J., and Neyer, F. J. (2017). Personality development in close relationships. In *Personality development across the lifespan*, pages 357–369. Elsevier.

Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, 89:122–135.

Fortin, B. and Yazbeck, M. (2015). Peer effects, fast food consumption and adolescent weight gain. *Journal of health economics*, 42:125–138.

Glaeser, E. L., Sacerdote, B. I., and Scheinkman, J. A. (2003). The social multiplier. *Journal of the European Economic Association*, 1(2-3):345–353.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, 48(1):26.

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

Golsteyn, B., Non, A., and Zölitz, U. (2017). The impact of peer personality on academic achievement. *University of Zurich, Department of Economics, Working Paper*, (269).

Graham, B. S. (2014). An empirical model of network formation: detecting homophily when agents are heterogenous.

Gray-Little, B., Williams, V. S., and Hancock, T. D. (1997). An item response theory analysis of the rosenberg self-esteem scale. *Personality and social psychology bulletin*, 23(5):443–451.

Harris, J. R. (1995). Where is the child's environment? a group socialization theory of development. *Psychological review*, 102(3):458.

Hartup, W. W. and Stevens, N. (1997). Friendships and adaptation in the life course. *Psychological bulletin*, 121(3):355.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Hogan, J. and Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of applied psychology*, 88(1):100.

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

Hoxby, C. M. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Citeseer.

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

Hsieh, C.-S. and Van Kippersluis, H. (2018). Smoking initiation: Peers and personality. *Quantitative Economics*, 9(2):825–863.

Ilmarinen, V.-J., Vainikainen, M.-P., Verkasalo, M., and Lönnqvist, J.-E. (2019). Peer sociometric status and personality development from middle childhood to preadolescence. *European Journal of Personality*, 33(5):606–626.

John, O. P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.

Lavy, V. and Sand, E. (2019). The effect of social networks on students' academic and non-cognitive behavioural outcomes: Evidence from conditional random assignment of friends in school. *The Economic Journal*, 129(617):439–480.

Lee, L.-f. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4):307–335.

Lee, L.-f., Liu, X., and Lin, X. (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal*, 13(2):145–176.

LeSage, J. P. and Pace, R. K. (2010). Spatial econometric models. In *Handbook of applied spatial analysis*, pages 355–376. Springer.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860.

Lin, X. (2015). Utilizing spatial autoregressive models to identify peer effects among adolescents. *Empirical Economics*, 49(3):929–960.

Liu, X., Patacchini, E., and Rainone, E. (2017). Peer effects in bedtime decisions among adolescents: a social network model with sampled data. *The Econometrics Journal*, 3(20):S103–S125.

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

Merlino, L. P., Steinhardt, M. F., and Wren-Lewis, L. (2019). More than just friends? school peers and adult interracial relationships. *Journal of Labor Economics*, 37(3):663–713.

Mund, M., Jeronimus, B. F., and Neyer, F. J. (2018). Personality and social relationships: As thick as thieves. In *Personality and disease*, pages 153–183. Elsevier.

Mund, M. and Neyer, F. J. (2014). Treating personality-relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology*, 107(2):352.

Mund, M. and Neyer, F. J. (2020). Wild horses dressed like unicorns: Relationship effects on personality. *The handbook of personality dynamics and processes. San Diego: Elsevier.*

Murphy, R. and Weinhardt, F. (2018). Top of the class: The importance of ordinal rank. Technical report, National Bureau of Economic Research.

Norris, J. L. (2018). *Identity Economics and Malleable Characteristics among Adolescents: A Study of Family and Peer Effects for Schooling Attitudes and Performance*. The University of North Carolina at Greensboro.

Nyhus, E. K. and Pons, E. (2005). The effects of personality on earnings. *Journal of economic psychology*, 26(3):363–384.

Patacchini, E., Rainone, E., and Zenou, Y. (2016). Heterogeneous peer effects in education. *Bank of Italy Temi di Discussione (Working Paper) No*, 1048.

Patacchini, E., Rainone, E., and Zenou, Y. (2017). Heterogeneous peer effects in education. *Journal of Economic Behavior & Organization*, 134:190–227.

Patacchini, E. and Zenou, Y. (2012). Juvenile delinquency and conformism. *The Journal of Law, Economics, & Organization*, 28(1):1–31.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2):322.

Reniers, R. L., Beavan, A., Keogan, L., Furneaux, A., Mayhew, S., and Wood, S. J. (2017). Is it all in the reward? peers influence risk-taking behaviour in young adulthood. *British Journal of Psychology*, 108(2):276–295.

Roberts, B. W., Walton, K. E., and Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin*, 132(1):1.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2):681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement.*

Santos, P. and Barrett, C. B. (2010). Identity, interest and information search in a dynamic rural economy. *World Development*, 38(12):1788–1796.

Störmer, S. and Fahr, R. (2013). Individual determinants of work attendance: Evidence on the role of personality. *Applied Economics*, 45(19):2863–2875.

Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6):1273–1296.

Udry, C. R. and Conley, T. G. (2004). Social networks in ghana. *The Social Economics of Poverty*, page 232.

van Zalk, M. H., Nestler, S., Geukes, K., Hutteman, R., and Back, M. D. (2020). The codevelopment of extraversion and friendships: Bonding and behavioral interaction mechanisms in friendship networks. *Journal of Personality and Social Psychology*, 118(6):1269.

VandenBos, G. R. (2007). *APA dictionary of psychology.* American Psychological Association.

Young, J. K. and Beaujean, A. A. (2011). Measuring personality in wave i of the national longitudinal study of adolescent health. *Frontiers in psychology*, 2:158.

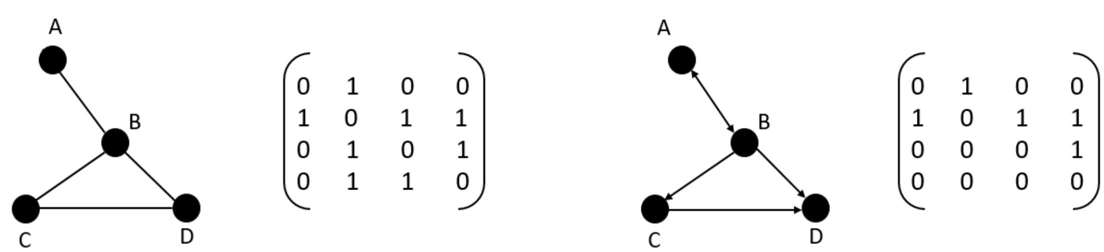
Zárate, R. A. (2019). Social and cognitive peer effects: Experimental evidence from selective high schools in peru.

Appendix I: Adjacency matrices

A graph, (N, E) is a collection of nodes, $N = \{1, ..., n\}$, and links (edges), E, between them. A graph can be represented as a matrix. An adjacency matrix $\mathbf{G}$, also known as an interaction matrix, represents who is whose neighbour in the social network. Each entry in the matrix $\mathbf{G}$ is given by g_{ij}, where:

$$g_{ij} = \begin{cases} 1, & \text{if } (i,j) \text{ is a link} \\ 0, & \text{if } (i,j) \text{ is not a link} \end{cases}$$

The figure below depicts a graph and the corresponding adjacency matrix. Panel I shows an undirected graph. The edges thus indicate a two-way relationship. Note that the adjacency matrix is symmetric in the case of an undirected graph. Friends on, say, Facebook can be represented using a undirected graph Panel II shows a directed graph i.e. the edges indicate a one-way relationship. To illustrate, followers on instagram can be represented using a directed graph.



Panel I Panel II

Graphs and adjacency matrices

In this paper, I employ the powers of the adjacency matrix. It is instructive to note that entry (i, j) in $\mathbf{G^s}$ gives the number of walks from i to j of length s. Thus, $\mathbf{G^2}$ gives us the adjacency for peers of peers and so on.

Appendix II: Description of data

Variable	Description
White	Takes value 1 if individual is white and 0 otherwise
Female	Takes value 1 if individual is female and 0 otherwise
Age	Age of the individual
GPA	Mean school performance based on individual's scores in English, social science, mathematics, and science. The scores are coded as 1 = D, 2 = C, 3 = B and 4 = A
Residential building quality	Interviewer remark to "How well kept is the building in which the respondent lives", coded as 1 = very poorly kept, 2 = poorly kept, 3 = fairly well kept, 4 = very well kept.
Household size	Household size of the individual
Parental education	Maximum parental education attained by either of the parents. "Never went to school", "not graduated from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college" are coded as 1–5.
Lives with mother	Takes the value 1 if the individual lives with their biological mother
Smokes or drinks	Takes the value 1 if the individual smokes regularly or has consumed alcohol outside supervision
Grade dummies	Dummies for grade of the student- grades 7-8, 9-10 and 11-12
Parent occupation dummies	Dummies for description of parents' jobs. If both parents are in the household, the occupation of father is considered. The categories are manager, professional/technical, office or sales worker, manual and other
Sports dummies	Two dummies for soccer/basketball/baseball/volleyball and other sports.
Extracurricular dummies	Dummies for the following groups: foreign languages, debate and drama, academic, music and dance, honour society/journal, other clubs

Appendix III: Extended main results

	Emotional Stability		*Extraversion*		*Conscientiousness*	
	2SLS (1)	ML (2)	2SLS (3)	ML (4)	2SLS (5)	ML (6)
Female	0.174***	0.168***	0.053	0.041	-0.008	-0.008
	[0.041]	[0.039]	[0.037]	[0.035]	[0.039]	[0.039]
Age	-0.009	-0.002	-0.006	-0.000	-0.057**	-0.057**
	[0.026]	[0.024]	[0.023]	[0.021]	[0.025]	[0.024]
White	-0.072	-0.063	0.007	0.012	0.054	0.050
	[0.069]	[0.065]	[0.062]	[0.059]	[0.067]	[0.067]
School performance (GPA)	-0.030	-0.039	-0.032	-0.043*	-0.075***	-0.078***
	[0.028]	[0.026]	[0.025]	[0.023]	[0.027]	[0.027]
Residence building quality	-0.066	-0.067	-0.041	-0.052	-0.001	-0.001
	[0.043]	[0.042]	[0.038]	[0.037]	[0.041]	[0.041]
Household size	0.035**	0.035**	-0.012	-0.010	0.025	0.024
	[0.017]	[0.017]	[0.015]	[0.015]	[0.017]	[0.017]
Parental education	-0.017	-0.015	-0.029	-0.030	-0.062**	-0.064**
	[0.027]	[0.026]	[0.024]	[0.024]	[0.026]	[0.026]
Lives with mother	-0.103	-0.100	-0.037	-0.059	0.000	-0.001
	[0.071]	[0.069]	[0.064]	[0.062]	[0.068]	[0.068]
Smokes or drinks	0.203***	0.198***	0.093***	0.081**	0.164***	0.163***
	[0.040]	[0.038]	[0.035]	[0.034]	[0.038]	[0.039]
Proportion female friends	-0.111*	-0.093	0.039	0.043	-0.008	-0.007
	[0.062]	[0.061]	[0.056]	[0.054]	[0.060]	[0.060]
Peers' average age	-0.053	-0.055	-0.031	-0.020	-0.010	-0.010
	[0.039]	[0.038]	[0.035]	[0.034]	[0.037]	[0.037]
Proportion white peers	-0.007	-0.017	-0.054	-0.064	0.147	0.154
	[0.108]	[0.105]	[0.097]	[0.094]	[0.104]	[0.105]
Peers' average GPA	-0.041	-0.039	0.016	0.019	-0.022	-0.021
	[0.043]	[0.042]	[0.038]	[0.037]	[0.041]	[0.041]
Peers' average residential building quality	0.062	0.078	0.052	0.057	0.040	0.047
	[0.066]	[0.065]	[0.059]	[0.058]	[0.064]	[0.064]
Peers' average household size	-0.009	-0.012	-0.001	-0.011	-0.003	-0.003
	[0.027]	[0.026]	[0.024]	[0.023]	[0.026]	[0.026]
Peers' average parental education	0.075*	0.074*	-0.059	-0.048	0.051	0.052
	[0.042]	[0.041]	[0.038]	[0.037]	[0.041]	[0.041]
Proportion peers live with mother	-0.062	-0.113	0.020	0.019	-0.066	-0.070
	[0.109]	[0.106]	[0.097]	[0.095]	[0.105]	[0.105]
Proportion peers smoke or drink	-0.035	-0.014	0.122**	0.129**	-0.018	-0.015
	[0.061]	[0.059]	[0.054]	[0.053]	[0.058]	[0.058]
Constant	-0.001	-0.001	0.025	0.015	-0.001	-0.002
	[0.018]	[0.015]	[0.017]	[0.014]	[0.018]	[0.017]
Endogenous peer effect	0.288***	0.292***	0.344***	0.353***	0.111	0.059
	[0.074]	[0.062]	[0.065]	[0.048]	[0.078]	[0.100]
ρ		-0.217***		-0.213***		-0.018
		[0.070]		[0.057]		[0.105]
N	2308	2308	2308	2308	2308	2308
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade dummies	Yes	Yes	Yes	Yes	Yes	Yes
Parent occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Extracurricular dummies	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors (in parentheses).
* p¡0.05, ** p¡0.01, *** p¡0.001