School ties: An analysis of homophily in an adolescent friendship network.

Simon Burgess, Eleanor Sanderson
and Marcela Umaña-Aponte

August 2011

Working Paper No. 11/267

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

School ties: An analysis of homophily in an adolescent friendship network.

Simon Burgess[1], Eleanor Sanderson[2]
and Marcela Umaña-Aponte[2]

[1]CMPO, *University of Bristol and CEPR*
[2] *CMPO, University of Bristol*

August 17, 2011

Abstract

Homophily is the tendency to establish relationships among people who share similar characteristics or attributes. This study presents evidence of homophilic behaviour for an adolescent friendship network of 6,961 links in the West of England. We control for unobserved characteristics by estimating school and individual fixed effects and present evidence on the role of length and closeness of friendships on the degree of homophily. We also exploit the dynamics of the friendship by comparing similarities among existing and future friends. Results indicate that academic achievement, personality, educational aspirations, bad behaviour and mother's education are essential in the friendship formation process. However, income and parents' occupational class proved to be insignificant. We also show that the degree of homophily among friends selected from a random process is much lower than that of the observed friendships.

Keywords: Networks, Homophily, Segregation, Friendships, Adolescents

JEL Classification: L14, C33, D83, Z13

Electronic version: www.bristol.ac.uk/cmpo/publications/papers/2011/wp267.pdf

Address for correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
Simon.burgess@bristol.ac.uk
www.bristol.ac.uk/cmpo/

1 Introduction

Social networks are ubiquitous and powerful. As Jackson (2010) says: "The people with whom we interact... influence our beliefs, decisions and behaviours" (p.1). The manner in which networks carry this influence depends in detail on the structure and characteristics of the network (see Jackson; 2010, for a thorough survey), and one prominent characteristic is the degree of homophily in the network. Homophily (first defined by Lazarsfeld and Merton; 1954) means a tendency of individuals to form links disproportionately with others like themselves. The degree of homophily in a network affects the speed of contagion across a network - for example, the spread of beliefs or behaviours. Golub and Jackson (2010) show that homophily slows down the speed at which a society reaches a global consensus. Therefore, they argue that understanding homophily is crucial to understand the functioning of a society. The degree of homophily can matter when agents' decisions are complementary, such as investment decisions (Calvo-Armengol and Jackson; 2009) or labour force decisions (Jackson; 2007), and generosity between agents (Goeree et al.; 2010). Homophily is also closely linked to the ideas of separation and segregation. High levels of homophily imply high segregation.

In our context of a large friendship network of adolescents, these effects of homophily seem particularly important. These individuals are making a transition between childhood and adulthood and their emerging attitudes and beliefs will be affected by their friendships. More transiently but of great practical importance is the spread of information and views around the network, for example, about risky behaviours.

We add to the network literature by exploiting a new and unique dataset. This is an adolescent friendship network of 6,961 links from the area around Bristol in England. There are numerous advantages to this data. It is a longitudinal dataset that has been collected since the individuals were born. Therefore, we have several measures of their academic achievement, ability, personality, behaviour, aspirations and socioeconomic status of their families. It also contains detailed information of the relationship and interaction of friends, such as, whether they were going to the same school, the place where they met, the length of their friendship, how much time they spend together and their tastes, activities and conversation topics. The dataset also has disadvantages, principally because its construction, as part of the ALSPAC study, means that not everyone in the network is a nominator (see below for details).

This is the first study using this data. Our aim in this paper is to characterise the degree of homophily in this network. We consider different dimensions of homophily, including both status and value homophily. We are also able to analyse the dynamics of homophily as we observe network members before and whilst they have friendship links. We have data on the links themselves, in addition to the characteristics of the nodes (the people); for example,

the length of the friendship, how close a friendship it is, and the bases of the friendship (for example, what their shared interests are). We can also control for unobserved characteristics at the individual level as some measures were taken over time.

We find that adolescents are very similar to their friends in their academic achievement (especially in KS2 and KS3 scores), total IQ, likelihood of going to university, popularity, bad teen behaviour (arriving late to and skipping classes) and their mother's education. In addition, some dimensions of personality such as extraversion, locus of control and intensity seeking are very important in the friendships formation process for teenagers. Family income and parents' occupational class do not seem to be relevant characteristics to establish friendships. So, one perhaps surprising finding is that socio-economic status homophily is relatively low.

Similarities in academic achievement, total IQ, social skills, popularity, personality, bad teen behaviour, likelihood of going to university and intensity seeking are stronger for close friends. Recently-met friends are less similar to respondents in some dimensions (academic achievement, IQ, playing musical instrument and extraversion, emotional stability, intellect/imagination) compared to whole-life friends. It is reasonable to think that very long friendships in adolescence (longer than 13 years) are very solid, and similarities (homophily) in personality between those friends should be stronger. Moreover, those friends could have influenced each other and shaped their personality together.

The rest of the paper is organised as follows. The next section reviews the relevant literature on social networks and homophily, and summarises the results from these studies. Section 3 discusses our dataset, the variables we use in the analysis and some sample selection issues. In Section 4 we describe the methodology. Section 5 presents the results, and Section 6 concludes.

2 Background

The social networks literature has identified a tendency for people to establish relationships with other people who share their same characteristics (Jackson; 2008, p.68-69). This behaviour is called *homophily*, and it is important to understand high levels of social segregation (Moody; 2001; Currarini et al.; 2009), the dynamics of labour market (Calvo-Armengol and Jackson; 2004), criminal (Haynie; 2002) and risk behavior (Kandel; 1978; Prinstein et al.; 2001; Lundborg; 2006; Pearson et al.; 2006), conflicts (Centola et al.; 2007), and the transmission of information (Rogers and Bhowmik; 1970; Golub and Jackson; 2009; Galeotti et al.; 2009) or disease (Bearman et al.; 2004; Rostila; 2010). People tend to bond with similars because communicating and developing closer ties, solidarity and trust with them might be easier. In addition, the cost of maintaining those ties is smaller.

Lazarsfeld and Merton (1954) classified homophily in two groups, status and value homophily. Status homophily is based on intrinsic characteristics of individuals such as ethnicity, age, gender and religion or acquired characteristics like education or occupation. In contrast, value homophily is related to internal states such as personality and future expectations, attitudes and aspirations (McPherson et al.; 2001).

People create links or relationships with others who are similar to them either by *choice* or by location (*induced* homophily). Location plays a key role on determining the possibilities for links creation. Therefore, the degree of homophily depends on the distribution of the characteristic in the (local) population, as this distribution determines the meeting opportunities of agents. In this context, homophily can be classified as: (i) *baseline* homophily if the proportion of friends who share similar characteristics is the same as the population; (ii) *inbreeding* homophily when the proportion of relationships that people form with people from their same type is greater than the observed in the population, (this is the most common occurrence); and (iii) *heterophily* when the proportion of relationships with people from other types (different from their own) is greater than the population distribution (Currarini et al.; 2009; Currarini and Vega-Redondo; 2010)[1].

A vast literature has evaluated status homophily[2], and especially, racial homophily. Currarini et al. (2009) and Currarini and Vega-Redondo (2010) find strong racial inbreeding homophily in a friendships network, using the National Longitudinal Adolescent Health survey (Add Health)[3,4]. Other recent studies on racial and ethnic homophily in in the United States (USA) are Fong and Isajiw (2000), Moody (2001), Mollica et al. (2003), Kao and Joyner (2004) and Doyle and Kao (2007).

---

[1]Homophily can be measured as the ratio of type g friends of a type g individual (i) to the total friends of i for each type-group. Therefore,

$$H_{i_g} = \frac{s_{i_g}}{t_{i_g}} = \frac{s_{i_g}}{s_{i_g} + d_{i_{-g}}}$$

is the homophily index of each type g individual which is then average over all type g individuals to get a final $H_g = \frac{H_{i_g}}{N_g}$. This is a simple index which is normally compared to the proportion of type g individuals in the population $\omega_g = \frac{N_g}{N}$

Baseline homophily occurs if $H_g = \omega_g$. However, when individuals are biased to make friends of their same type, the homophily index is greater than the population proportion $H_g > \omega_g$, this is called *inbreeding homophily*. On the other hand *heterophily* arises if $H_g < \omega_g$.

[2]see McPherson et al. (2001) for an overview.

[3]Add Health is an adolescent friendship network which surveyed a representative sample of adolescents who where in grades seven to twelve in 145 schools spread across 80 different areas in the United States during 1994-95. 90,118 students filled out the first wave of the questionnaire which was completed at school. In this questionnaire respondents were asked to name up to 5 male and 5 female friends, and in almost all cases students nominated fewer than 5 friends for each gender. For friends that went to the same school as the respondent, these nominations where then matched to the person they had nominated to create a friendship network for the school, however no information was available for the 15% of nominated friends that did not go to the same school as the respondent. There were three additional waves in 1996, 2001-2002 and 2008 in which respondents were interviewed at their homes.

[4]Currarini and Vega-Redondo (2010) also find inbreeding homophily in marriage markets using a USA census data from IPUMS.

In addition to racial homophily, Shrum et al. (1988) explore gender homophily in a sample of friends from an American school (junior, middle and high school). They find that racial homophily increases and gender homophily decreases with school grade. Baerveldt et al. (2004) also look at gender and ethnic homophily for a sample of Dutch adolescents (16-18 years old) from 20 urban high schools. Their findings suggest a high tendency to gender homophily (especially for girls) and ethnic homophily in all four ethnic groups (Dutch, Moroccan, Turkish and Surinamese) although less strong for Dutch individuals, contradicting the traditional belief that natives (normally the majority group) are less willing to have inter-ethnic relationships. Wimmer and Lewis (2010) compare the magnitude of racial homophily with other characteristics (ethnic and micro-ethnic, regional origin, socioeconomic status, cultural tastes, academic major and co-residence) in a sample of friends at an American university. They show that race is important to create social ties but it is not the dominant factor, as other characteristics such as sharing a dorm room or doing the same academic major at the university are equally or more relevant. Hence, they emphasise the importance of evaluating homophily in various dimensions. Additionally, Marmaros and Sacerdote (2006) argue that "geographic proximity" is a crucial determinant of link formation among college students, as the probability of inter-racial relationships increases, on average, by a factor of three when black and white students share a room at a dorm. Other factors, such as academic achievement (measured as high school GPA), might also play a key and stronger role in friendship formation compared to region of origin, class year or race (Foster; 2005)[5]. Kossinets and Watts (2009) construct a *similarity index* using six status dimensions (gender, age, status, field, year, and country of origin - USA/foreigner) for a large university community. They show that similar pairs show a higher propensity to form new ties.

Adult connections are also characterised by high levels of status homophily in gender, age, race, religion, education and occupation, in labour networks or confiding relations (McPherson and Smith-Lovin; 1987; Marsden; 1988). Recent studies have looked at the online dating market, finding that males and females normally contact and reply to people who have their same level of education[6] or who are similar in age and physical attractiveness, stating that dominant linking mechanism is education (Skopek et al.; 2010). Hitsch et al. (2010) look at additional characteristics and show that people with similar marital status, fertility[7], looks, BMI, education, ethnicity and religion are more likely to contact each other. They identify strong selective homophily in terms of education, ethnicity and religion, but not for income.

Few studies have investigated value homophily. Cohen (1977) compare similarities among

---

[5]She uses a sample of students from an American university and controls for unobserved heterogeneity with student fixed effects

[6]sometimes higher but never those with lower educational attainment.

[7]Therefore, single people avoid divorced people and divorced women prefer to contact divorced men. Members with children tend to contact other members with children, and they are less attractive to people who do not have children.

friends from an American school on risk behaviour (alcohol drinking and smoking), academic interest (taking extra courses, time spent on homework everyday), aspirations (value of learning as much as possible, intention to go to college, desire to be the leader in school activities or be remembered as a star athlete or brillian student) and good behaviour (value of pleasing parents and of living up to religious ideals and church attendance.). He finds that the members of 49 groups of close friends are very homogeneous in most of the studied characteristics, and observes that similarites are mainly the result of "homophilic selection" rather than of "conformity pressures or leaving of the group by deviates". Similarly, Kandel (1978) analyses risk behaviour (cannabis use and minor delinquent activities), level of academic aspirations and political identification in 957 pairs (dyads) of best friends from five New York high schools. She argues that friends share similar attributes due to both choice (homophily selection) and influence (peer effects) and that both factors play a very important role as homophily increases with constant interaction.

Recent work on value homophily (sometimes combined with status characteristics) shows that American and Indonesian children are similar to their friends in academic performance, achievement beliefs, motivational beliefs, social preferences and behavioural problems (Altermatt and Pomerantz; 2003; French et al.; 2003). In addition, American adolescents tend to share similar academic goals and achievements, risk behaviour (substance use), ethnic identity (Hamm; 2000) and suicidal thoughts and attempts (Bearman and Moody; 2004). Also, happiness clusters have been identified in a network of people composed of parents, siblings, children, friends, neighbors and acquaintances (the Framingham Heart Study). People who report being happy tend to be connected with other happy people and this happiness is spread through the network (Fowler and Christakis; 2008).

The most relevant study for our purposes is Bearman et al. (2004) in which the authors evaluate similarities in 22 dimensions (status and value) for an adolescent romantic network from one of the largest schools[8] in the Add Health project. In particular, they show that adolescents tend to select partners with similar socioeconomic status, academic performance (GPA), academic aspirations (likelihood of going to college), risk behaviour (drinking, criminality, smoking), popularity (number of nominations as a friend), IQ (Vocabulary test), religion and sexuality.

Our paper adds to the literature on social networks because it analyses the friendships formation process and homophily degree of a group of 16 year old secondary school children in the United Kingdom (UK). We evaluate status and value homophily in more than 28 individual characteristics from a novel dataset, as described in the following section.

---

[8]Almost all students in this school were white, as in our sample.

3   Data and Descriptive Statistics

3.1   ALSPAC cohort

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a long-term study which aims to evaluate the health and development of a cohort born in the Avon area of England. Initially, the study enrolled 14,541 pregnant women[9], whose expected date of delivery fell between 1st April 1991 and 31st December 1992. 13,801 (95%) of those women became the mothers of surviving offspring at 12 months, for a total of 13,971 children in the study at that age (including multiple births). The Avon area has a population of one million and includes the city of Bristol (population 0.5 million), and a mixture of rural areas, inner city deprivation, suburbs and moderate sized towns[10] (Gregg et al.; 2007).

Study families were surveyed with high frequency from the time of pregnancy onwards. Mothers completed four postal questionnaires prior to the birth, plus a further five on family characteristics and a further eight focusing on the study child in their first year of life. After this, parents, teachers and study children have responded 79 additional questionnaires about their development, experiences and the growth environment. The study also contains data from a number of other sources. Eight clinics took place when the children were 7 to 15-16 years old, in which children were administered a range of detailed hands-on physical, psychometric and psychological tests. A number of external sources of information have also been matched to the ALSPAC children. These include records from the National Pupil Database (NPD), which contains school identifiers and results on national Key Stage school tests for all children in the public school system, and information of local deprivation at the small area level (the government-produced Indices of Multiple Deprivation, IMD) (Gregg et al.; 2007).

3.2   Adolescent Friendship Network

A friendships questionnaire was sent to 7,865 study adolescents (15-17 years old) in March 2008 in which they were asked to nominate a maximum of five best friends. They were explicitly told that the nomination order was not relevant (i.e. the first nominated friend will not be taken as the best best friend). The exact full name, gender, ethnicity and date of birth (if known) of each friend were requested. Additional information about the interaction and relationship of the respondent and each friend were also collected, such as whether the

---

[9]Believed to be between 80 to 90 percent of all those who had a pregnancy during this period

[10]The 1991 census was used to compare the population of mothers with infants under one year of age resident in Avon with those in the whole of Britain. The sample is broadly representative of the national population although the mothers of infants in Avon were slightly more likely to be affluent, on average, than those in the rest of Britain. The ALSPAC sample is not entirely representative of all eligible mothers in the area, with a slight shortfall again in the less affluent and non-white mothers. See www.alspac.bris.ac.uk.

respondent and friend go to the same school, where they met, the length of their friendship, how much time they spend together at school (if applicable) and in school holidays, what kind of activities they do together outside school, what they talk about, whether friend shares same tastes of respondent, whether respondent have talked to friend about a problem in the previous week, whether friend is boyfriend/girlfriend and how they mostly communicate to each other.

The identification and matching of nominated friends is the central data task and was done in two stages. First, we attempted to identify the nominated friends using their (reported by the respondent) full name, date of birth and school. Gender and school year were also used but as there was not much variation in these variables they were less informative. Then, each nominated friend was compared to every other nominated friend in the dataset. Each resulting pair of friends was then given a score based on how similar they were in terms of first name, surname, day, month and year of birth and school. Using a book of names, full names, nicknames and abbreviations were standardised and compared to the name given. Therefore, where one of the names in the pair was a standard nickname for the other name in the pair, they will receive a higher score. This was done to account for where common abbreviations of names do not share many letters with the original name. Additionally, less common names were given a higher score than more common names as having the same uncommon name was more likely to indicate a match than having the same common name. The pairs where divided up into three groups based on the score they received, those that were a definite match (which had a high score), those that were definitely not a match (low score)and those that the score alone did not confirm whether the pair were a match or not. From the third group a sub-set of names was randomly selected and then manually inspected to determine if it was a match or not, and thus, to find the cut-off score between non matches and matches. In order to make this process accurate enough and determine exactly where the boundaries lay, a further sub-set was randomly selected but from a much narrower range of scores around the level where some names were matches and some were not, until a clear boundary was selected. In the second stage, the resulting list of friends from the matching process was then checked against the information for the people in ALSPAC, so that any nominated friends who where in ALSPAC could be linked to their data, including their response to the friendship questionnaire in case they had responded.

A total of 3,123 (39.7%) respondents completed the questionnaire and each nominated on average 4.65 friends. The final data set contains a total of 14,503 friendship links, to 11,041 friends. There are, in total, 13,046 individuals in the sample as 1,404 (45%) respondents were also nominated as a friend. Due to the survey design, not all nominated friends are part of the ALSPAC study, thus, 40% (4,572) of friends are ALSPAC children[11]. In addition,

---

[11]This figure includes the respondents who were nominated as friends.

not all links are observed as not all study participants answered the questionaire. Therefore, we cannot observe the complete network in which "all possible links are present" (Jackson; 2008). Our network consists of 2,396 respondents and 6,961 links, i.e. 2.9 nominated friends per respondent on average. 80% of respondents in our sample nominated at least two friends; they represent 93% of the friendship links (see Table 1 for details on the frequency of friends nominations).

3.3 Outcome measures from ALSPAC

We have information on the academic performance, physical development and psychometric and psychological characteristics of 5,633 study teenagers[12]. The variables used in the analysis are:

Academic achievement. We use as measure of achievement the results of the national tests administered to all school children in the public sector. They are the Key Stage 1 (KS1) assessment in Year 2 at age 7, the Key Stage 2 (KS2) assessment in Year 6 at age 11 and the Key Stage 3 (KS3) assessment in Year 9 at age 14. Each assessment tests pupils attainment in English, mathematics and science (Key Stage 2 and 3 only). We compute the average of the subject-specific scores to create an overall score[13], which is then standardised.

Intelligence Quotient (IQ). We use two measures of IQ. The first one is the short form of the Wechsler Intelligence Scale for Children (WISC-III UK), administered by ALSPACs psychology team to children at the age of 8[14]. We compute a raw total IQ score, by adding the individual scores on five verbal sub-tests: information, similarities, arithmetic, vocabulary and comprehension; and five performance sub-tests: picture completion, coding, picture arrangement, block design and object assembly. The second one is the Wechsler abbreviated scale of Intelligence - WASI (Psychological Corporation, 1999) administered by ALSPACs psychology team to children at the age of 16. The raw total IQ score is also the sum of individual scores in two sub-tests: word vocabulary (28 questions) and matrix reasoning (29 questions). Both measures are then standardised to make them comparable.

Self esteem. School achievement, expectations for success and social relationships might be affected by how much people believe and are confident with themselves (e.g. see Damon & Hart, 1982). Self esteem was measured at age 8 at a clinic using a 12-item shortened form of Harters Self Perception Profile for Children (Harter, 1985) comprising the global self-worth and scholastic competence subscales. Responses to the 12 items were scored such that a higher score indicates higher self esteem.

---

[12]Respondents (2,396) and friends (3,237).

[13]For the Key Stage 2 and 3 assessments we construct a finer measure for each subject using additional information on pupils' marks before averaging over the three sub-scores.

[14]This was then the most up-to-date version of the WISC, the most widely used individual ability test world-wide. See Wechsler, Golombok and Rust, (1992)

Child behavioural problems. We use the mother's responses to the Strengths and Difficulties Questionnaire - SDQ (Goodman, 1997) when the child was 7, 12 and 13 years old. The SDQ is a behavioural screening questionnaire that comprises 25 questions relating to five dimensions of behaviour pro-social, hyperactivity, emotional symptoms, conduct problems and peer problems[15] . Respondents are asked to indicate the extent to which 25 statements matched the study childs behaviour over the last six months. A maximum of ten can be scored for each component. We sum four of the sub-scores, excluding the pro-social score, to obtain the total difficulties score. A high score indicates abnormal behaviour.

Teen behavioural problems. Adolescents reported the frequency (never, sometimes, often) they use to arrive late to lessons or skip classes at the age of 15.

Future expectations and aspirations. Measured at 15 years old as the likelihood of going to university in the following five years after finishing the compulsory school.

Personality:

- Five Factor Model (FFM). The five-factor model (FFM) identifies "big five" relatively independent, broad dimensions of personality: extraversion, agreeableness, conscientiousness, emotional stability and culture, intellect, or openness to new experience. The International Personality Item Pool (IPIP) was employed to investigate the Big Five personalities at 13-14 years old. The measure consist of 50 statements (see appendix B) which are rated on a five point scale[16]. Some of the statements were positive keyed and some negative keyed[17]. For the negative keyed statements, the values are reversed. The total score is the sum of the assigned values for each item.

- Arnett's Inventory for Sensation Seeking (AISS). This measure evaluates two personality dimensions: (i) the need for varied, novel and complex sensations/experiences (*novelty*) and (ii) the willingness to take physical and social risks (*intensity*). It has been used in studies on potential risk behaviour (The ALSPAC Study Team; 2010). Adolescents (17 years old) were asked to answer 20 questions (see appendix C)., each with a score from 0 to 4[18]. The total score is the sum of scores for each dimension. AISS was also applied in clinics at 11 and 14 years old. However, only 11 questions of the original version of the AISS were chosen for inclusion and a further 9 questions were re-designed to make them more age-appropriate than the original questions. The majority of modified questions were related to novelty seeking, for this reason they are not comparable to the 17-year old measure and will not be used in the homophily analysis.

---

[15]This measure is a good predictor of conduct, emotional, hyperactivity and any psychiatric disorders in children of the age we examine (Goodman et al., 2000).

[16]The adolescent selects among "Very like me", "Quite like me", "Neither like me or not like me", "Not much like me", "Not like me at all".

[17]http://ipip.ori.org/newScoringInstructions.htm

[18]Describes me very well=4, Describes me a bit=3, Does not describe me very well=2, Does not describe me at all=1 for the positive statements and the values are reversed for the negative statements.

- Locus of control. It is a measure of the perception of a connection between one's actions and their consequences. People who believe that an outcome is largely the result of their own behaviour or actions are seen as having a more internal locus of control, whereas individuals with an external locus of control tend to attribute outcomes to luck, chance, fate or the interventions of others. Locus of control appears to be an important factor in the choices people make. Internal individuals are expected to be more active in pursuing goals and to show greater ingenuity and persistence when confronted with obstacles than external individuals. Our measure of locus of control is the shortened version of the Nowicki-Strickland Internal-External scale (NSIE scales). The scale consists of 12 questions (see appendix D) each requiring a yes/no answer. The questions were read out to the child by an examiner during an ALSPAC clinic at age 8 and asked on a questionnaire sent to ALSPAC young persons (17 years old). Each response was coded 0 or 1 and added to create a total score.

Friendship variables. During two clinics (at 8 and 10 years old) children were asked five questions based on the Cambridge Hormones and Moods project Friendship questionnaire (Goodyer et al, 1989, 1990) to measure how happy they were with their friends and the quality of their friendship (see Appendix A). A *friends score* was created by recoding the answers such that a higher score denotes the most positive state. Children who responded "don't know" to at least one question were excluded from the score, as were those with at least one missing response. Children also reported how many close friends they had at 10, 11 and 13 years old. At nine years old they were asked whether they had lots of friends, made friends easily or had more friends than most other kids. Parents also reported whether their children talked about school friends or whether they tend to be solitary at 11 and 13 years old.

Family socio-economic status. Measures of socio-economic status are parents' occupational class and parents' highest educational qualification reported when the child was born; housing tenure when the child was 8, 21, 33 and 61 months old; the log of the average of equivalised net household income at ages 33 and 47 months, expressed in June 1995 prices[19] and a financial difficulties score, which was calculated from the mother's answers to whether she has had difficulty affording food, clothing, heating, rent/mortgage and items for their child when the child was 8, 21, 33 and 61 months old.

Anthropometric Measures. *Height* - height was directly measured eight times, in all clinics, from age 7 to 12, 14 and 16. *Fat mass* - The fat mass of children was also directly

---

[19]Income data from the ALSPAC data is banded. A median value for each band using data from the Family Expenditure Survey was imputed to convert the income variables to real values using the 1995 RPI as a base, and then equivalise using the OECD modified scale. We also impute the value of housing benefit for families who do not directly receive housing payments. Finally, we average over the two measures to reduce measurement error and take the log of the variable.

measured at age 9, using a dual-energy X-ray absorptiometer (DXA), a costly method involving a full body scan that is highly accurate (Morrison et al., 1994). Total body fat mass is measured in grams adjusted for age of child in months, sex, height and height squared[20]. *body fat percentage* was measured at 11, 12, 13 and 15 years old. These were also adjusted for age of child in months, sex, height and height squared[21].

Section 3.5 presents descriptive statistics of the data and discuses similarities and differences between respondents and non-respondents of the friendships questionnaire.

3.4   Closeness measure

The friendships questionnaire collected information about how much time respondents and friends spend together at school[22] and during school holidays, the topics they talk about and activities they do together[23], and if they talk about problems. Respondents also had to report if they knew their friends' date of birth. These questions were used to create a measure of closeness by assigning a value to each of them (see Appendix E with questions and their associated score). Friendships were classified by gender composition[24] and whether the friend goes to same school as respondent. Therefore, six groups were created, and a total value calculated for each group. Table 2 details the groups and contains descriptive statistics of the total score for each of them.

The total score is, then, standardised[25] for each group, and from them we create a single closeness index that measures of how close respondent and friends are at 15-16 years old. Figure 1 plots the distribution of this index by nomination order and Table 3 provide basic descriptive statistics.

3.5   Sample selection

As the response rate of the friendship questionnaire was 40%, we compare individual characteristics of respondents and non-respondents using information collected in previous surveys. Table 4 shows this comparison and presents a t-test of sample means for both groups (column 4). A higher proportion of girls and white participants answered the questionnaire. Non-respondents seem to be taller, but this might be explained by the fact that the proportion

---

[20]Obesity is defined as an excess of body fat. The measure most commonly used to define obesity is body mass index (BMI), the ratio of weight to height squared, which is a measure of over/underweight rather than of lean/fat. Although the two are highly correlated, individuals who are unusually muscular may be overweight but not fat, and hence screen false positive for fatness according to BMI (Power et al., 1997).

[21]Body fat percentage was not measured using the DXA and is not comparable with total body fat mass

[22]Note that if friend is not in same school as respondent this question is not applicable.

[23]Kao and Joyner (2004) show that shared activities is an indicator of frienship intimacy.

[24]Gender is important as topics/interests may differ depending on the sex of friends.

[25]$Z = (X_{ig} - \bar{X}_g)/sd(X_g)$ in which g=gender-school group, X_i= closeness score of friendship i, and $\bar{X}_g$ and $sd(X_g)$ are the group-specific mean and standard deviation of the closeness score (X).

of female respondents is higher. Respondents seem to be heavier but this does not translate into a higher obesity rate[26]; they perform better at school, their intelligence quotient (IQ) is higher and they show less behavioural problems than non-respondents. However, the self esteem score of both groups is similar. In terms of personality, there are no differences in the level of extraversion, emotional stability or intensity seeking, but respondents seem to care more about people (agreeableness), show more awareness of their surrounding world (conscientiousness) and are more open to new ideas or situations (openness). According to the locus of control score, respondents seem to be more internal-type. Finally, non-respondents show more interest on exploring or discovering unknown things, places or people according to the novelty score of the Arnetts Inventory for Sensation Seeking.

Table 5 explores the differences in terms of friendships at different ages. Respondents report a slightly smaller number of close friends in adolescence (11 and 13 years old). However, their friends score in childhood (8 and 10) is similar. Respondents and non-respondents seem to make friends equally easy and both report to have lots of friends at 9 years old, but respondents did not considered they had more friends than most other kids. Respondents seem to talk more about their friends with their parents. However, we do not find statistical differences in the social/friendship behaviour reported by their parents and measured as their tendency to be lonely.

Individual characteristics of parents were also explored. Table 6 indicates that parents of respondents are more educated, belong to higher occupational class and have a better socio-economic position with less financial difficulties.

4   Methodology

The standard measures of homophily are typically based on the relative likelihood of own-group links to all links. For example, Jackson (2010) and Currarini et al. (2009) who define the degree of homophily in a network of many groups. They also define "in-breeding" homophily as the case where links are more likely to be formed within groups than across groups. This type of definition is not so useful when applied to our data as we observe only a subset of links since not all agents are nominators (see Section 3.2).

We therefore approach the analysis of homophily from a different perspective. We examine the degree to which friends are alike, over different domains and controlling for different factors. We also make use of our data on different characteristics of the links themselves, such as the length and the strength of the links, and see if these are associated with differences in homophily. We adopt two strategies to control for variations in the nature of pools of potential friends facing the respondents (details below).

---

[26]Obese are those children (aged 9) whose DXA measure fell at the top 10th decile of the distribution.

We adopt a regression framework to characterise the alikeness of friends. This has a number of advantages. It allows us to deal with all five friends together, to control for some basic characteristics of the respondent (age and gender), to introduce the link characteristics in a natural way and to include and exclude school fixed effects.

Let y_i be a characteristic of young person i; for example, the value of a test score. The simplest model would simply compare the test score of i with that of her friend, $j_{(i)}$: $y_i = \alpha + \beta\, y_{j_{(i)}}$. In this regression, a friendship link would be the unit of observation, and the coefficient β would measure the average degree of alikeness of friends for that particular characteristic. This is our measure of homophily. We expand this simple model in a number of ways.

First, we add more friends and allow β to differ by friend,

$$y_i = \alpha + \sum_{j=1}^{5} \beta_j\, y_{j_{(i)}} + \varepsilon_i,$$

Then, we add respondents' gender ($female$), respondents' and friends' age in months (age_i and $age_{j_{(i)}}$), and the characteristics of the link itself interacted with the friend's characteristic,

$$
\begin{aligned}
y_i =& \alpha + \sum_{j=1}^{5} \beta_j\, y_{j_{(i)}} + \gamma_1\, female_i + \gamma_2\, age_i + \gamma_3\, \overline{age}_{j_{(i)}} \\
& + \gamma_g\, gendercomposition_{ij_{(i)}} \times y_{j_{(i)}} + \delta\, sameschool \times y_{j_{(i)}} + \varepsilon_i,
\end{aligned}
\tag{1}
$$

in which, $gendercomposition_{ij_{(i)}}$ are three categorical variables that identify the gender composition of the friendship (girl-girl, girl-boy, boy-girl[27]; $g = [4, 6]$). $sameschool$ indicates whether respondent and friend go to the same school.

Finally, we add the length of the friendship and the closeness measure (see Section 3.4) to evaluate their effect on the degree of alikeness. For simplicity, the model now has a pooled coefficient (β) for all friends[28]

$$
\begin{aligned}
y_i =& \alpha + \beta\, \overline{y}_{j_{(i)}} + \gamma_1\, female_i + \gamma_2\, age_i + \gamma_3\, \overline{age}_{j_{(i)}} + \gamma_g\, gendercomposition_{ij_{(i)}} \times y_{j_{(i)}} \\
& + \delta_1\, sameschool \times y_{j_{(i)}} + \delta_l\, length_l \times y_{j_{(i)}} + \delta_c\, closeness_c \times y_{j_{(i)}} + \varepsilon_i,
\end{aligned}
\tag{2}
$$

in which, $length$ are five dummy variables ($l = [2, 5]$) that group, in five categories, the length of the friendship: $length1$ - recent friend (0-2 years), $length2$ - friends met 3-4 years ago,

---

[27]The omitted category is boy-boy friendships.

[28]In section 5.2 we will test the equality of friends' coefficients and show that in most cases we cannot reject the null hypothesis of equality.

*length*3 - friends met 5 years ago (at the start of secondary school or end primary school), *length*4 - friends since primary school (6-12 years) and *length*5 - whole-life friends (before primary school - 13+ years)[29]. *closeness* are also five dummies, one for each quintile of the closeness distribution ($c = [6, 10]$). The first quintile contains the least close friends and the fifth quintile the closest ones. *length* and *closeness* are interacted with the outcome of friends. *sameschool* changes according to the time of the measure, i.e. under 11 years old it refers to "going to same primary school" and above 11 years old "going to same secondary school".

As it is clear from the literature (Currarini et al.; 2009; Jackson; 2010; Golub and Jackson; 2010), homophily can arise either from preferences for similar friends, or simply from the nature of the pool of potential friends. We adopt two approaches to deal with this. First, we include school fixed effects. This controls for all characteristics of the potential friends' pool in the school, including the relative frequencies of characteristics in the school-based friend pool, the nature of the "arena" in which potential friends meet and so on. School fixed effects are likely to be important for two reasons. First, we find that most friends (85%) come from the same school as the respondent, so within-school friends are crucial. Second, students are not randomly allocated to schools in England and there is considerable social sorting across schools. Thus the potential friend pool for a student in one school will be different in important ways to the potential friend pool for a student in another school. λ_s in Equation (3) characterizes school fixed effects.

$$
\begin{aligned}
y_{is} =& \alpha + \sum_{j=1}^{5} \beta_j \, y_{j_{(i)}s} + \gamma_1 \, female_{is} + \gamma_2 \, age_{is} + \gamma_3 \, \overline{age}_{j_{(i)}s} \\
& + \gamma_g \, gendercomposition_{ij_{(i)}s} \times y_{j_{(i)}s} + \delta \, sameschool \times y_{j_{(i)}s} + \lambda_s + \varepsilon_{is},
\end{aligned}
\tag{3}
$$

Our second approach to dealing with differential friend availability is to generate "simulated friends". We draw five[30] random "potential friends" for each respondent from the full sample of ALSPAC participants[31], who attended their same school and academic year. Using these simulated friendship links datasets, we estimate Equations (1) and (3)[32] 130 times. This is different to the fixed effects approach because we draw five simulated friends to mimic five real friends and control for sample size issues.

---

[29]The omitted variable is *length*5 - friends for 13+ years.

[30]Four of the same gender and one of the opposite gender.

[31]Excluding the respondent and their actual nominated friends.

[32]*sameschool* variable is excluded from the model as all simulated friends go to the same school as the respondent.

The dynamics of friendship and homophily We explore the dynamics of friendships by splitting the sample in two groups, friendships which already existed at the time of the measure (true friends) and those which would form afterwards (future friendships). Again, for simplicity, the model now has a pooled coefficient (β) for all friends. In addition, the variable *sameschool* changed to *sameprimschool* if the time of the measure was before 11 years old and children were already friends. It is a dummy variable indicating if both children were attending the same *primary* school[33]. We then add school fixed effects to the specification, and similarly, estimate primary school fixed effects for friendships formed before 11 years old.

Individual Fixed Effects Some outcomes were measured at different ages. We can use them to estimate individual (or respondent) fixed effects (IFE), as follows,

$$
\begin{aligned}
y_{it} =& \alpha + \beta\, \overline{y}_{j_{(i)}t} + \gamma_1\, age_{it} + \gamma_2\, \overline{age}_{j_{(i)}t} + \gamma_g\, gendercomposition_{ij_{(i)}} \times y_{j_{(i)}t} \\
& + \delta\, sameschool \times y_{j_{(i)}t} + \lambda_i + \eta_t + \varepsilon_{it},
\end{aligned}
\tag{4}
$$

in which λ_i and η_t are individual and year fixed effects, respectively. *sameschool* is a dummy variable which indicates if the respondent and friend have always gone to the same (primary or secondary) school. IFE allow us to control for all the unobservable characteristics of the respondent, like their own tastes, feelings, likes, dislikes, etc. We also add the closeness and length of friendship variables as in Equation 2 and the estimations are for true friendships only.

5 Results

5.1 Homophily: Graphical examination

As an initial examination of how similar the characteristics of respondents and friends are, Figures 2 to 7 present their relationship for a selection of status and value measures. For each characteristic, respondents were classified into ten quantiles of the distribution. Then, the mean of the same measure for all friends was plotted against the respondent's quantiles. This is done for existing friendships at the time of the measure. The figures show a high correlation among friends and respondents characteristics as the friends' means are normally higher the higher the respondent falls in the distribution. This indicates that respondents and friends seem to be similar (or homophilic), especially, in academic achievement, IQ and some aspects of their personality such as extraversion, inteligence/imagination, intensity

---

[33]Notice that for future friends the variable *sameschool* is kept, indicating whether they go to the same secondary school.

seeking and locus of control.

In the second column, we present analogous figures using measures for simulated friends. A sample of potential friends[34] was drawn from the full ALSPAC dataset and five potential friends were randomly selected. The curves of these simulated friends are, in general, flatter than of the real friends, indicating that the correlations among those randomly selected friends and the respondents are much lower than among real friends and respondents.

5.2   Homophily across different individual characteristics

Tables 7 to 10 present the results of equations (1) and (3), in which the outcomes are individual characteristics measured after 14 years old or parental characteristics[35]. We find strong correlations in parental education, income and occupational class, and academic achievement (KS3 score) (Columns 1-5). The coefficients of all five nominated friends are positive and significant for these characteristics. Some dimensions of personality also play and important role to develop close connections among adolescents. For example, extroverts tend to have extrovert friends. Also, individuals with high interest in adventures and intense sensations associate with each other. When we control for school fixed effects some of these strong connections disappear. In particular, most of parental characteristics become insignificant, as well as some dimensions of personality (locus of control, novelty and intensity seeking). However, academic achievement, mother's education and extraversion seem to be the three most important characteristics to form friendships.

Simulated friendships

We present the means of the 130 estimated coefficients of homophily and their standard errors for simulated friends (see Section 4), in columns 6-10 of Tables 7 and 9. The correlations in parental characteristics and academic achievement are positive and significant (OLS results), but fully fade away after adding school fixed effects. Also, note that the average degree of alikeness that we find (either significant or not) with the simulated data sets is much lower than the one estimated with the real friendship links. This result suggests that individuals tend to select their friends deliberately, and that, therefore, friendship link formation is not a simple random social selection process.

The coefficients on individual's and link's characteristics are in Tables 8 and 10. The alikeness in family income and conscientiousnes increase when the respondent is a girl. However, this effect vanishes in school fixed effects estimations. When one of the friends in the link is a girl, their KS3 results tend to be more similar as when respondent and friends go to the same school. However, this last result is the opposite for personality measures such

---

[34]those who go to the same school and are in the same academic year
[35]We consider these characteristics as unchanging.

as extraversion and locus of control (coefficients are negative and significant). This is an interesting result as it might indicate that the personality of friends who were met outside school is more alike.

Note that the magnitude of coefficients for all (nominated or simulated) friends A-E in Tables 7 and 9 is very similar. We test the equality of these coefficients and cannot reject the null hypothesis of equality for any characteristic, as shown by the F-tests results in the first and ninth columns of Tables 8 and 10. Therefore, hereafter all models will be estimated with a pooled coefficient for all nominated friends.

5.3 The dynamics of friendship and homophily

In this section we select characteristics measured in childhood or early adolescence to exploit the dynamics of the data. We and split the sample in two groups: (i) actual/existing friends at the time of the measure and (ii) friends met after the measure was taken (future friends). Table 11 presents the estimates of Equation (1), but with a single coefficient for all friends (β). We find positive and significant correlations in the academic achievement (KS1, KS2 and KS3 scores) of respondents and their actual friends. In particular, if the average KS3 score of all nominated friends is higher by 1 standard deviation, the respondent's KS3 score will be 0.383 standard deviations higher. We also find positive and significant correlations in existing friendships for total IQ, difficulties score, self esteem, playing a musical instrument, the likelihood of going to university and intensity seeking. Some of these correlations are smaller or disappear for friends met after the measure was taken. This is the case of total IQ, self-esteem and playing an instrument. However, future friends tend to be more similar in other characteristics such as locus of control, popularity in childhood[36], behavioural problems (difficulties score, arriving late to lessons or skipping classes), how likely they will go to university and intensity seeking. We do not find correlations in social skills (making friends easily or has lots of friends at 9 years old) or body fat.

By adding school fixed effects, we control for unobserved heterogeneity at the school level. Therefore, we are only comparing children who were exposed to the same environment and not across all children in our sample. Results show that children and their friends are still similar in their academic achievement (KS1, KS2 and KS3 scores), total IQ, likelihood of going to university and intensity seeking (Table 12). However, the magnitudes are smaller after controlling for school fixed effects. Similarities in self esteem, playing a musical instrument and difficulties score become insignificant, indicating that the school environment played an important role in the selection process of friends for these characteristics. Conversely, homophily in bad teen behaviour gets stronger (higher magnitudes) than in the simple OLS model, for existing friendships. Body fat is still not significant.

---

[36]Child has more friends than most other kids.

5.4 Factors affecting the degree of homophily

Now, we extend our analysis to evaluate how the length and closeness of the friendship affect the degree of homophily[37]. Results of Equation (2) indicate that similarities in academic achievement, total IQ, popularity, personality, bad teen behaviour, likelihood of going to university and intensity seeking are stronger the closer the friend is (Table 13). We find negative and significant correlations for friends who fall in the first quintiles (i.e. least close friends) for those measures. Consider now the big five personality dimensions. We find strong correlations with negative and significant coefficients in extraversion, agreeableness and intellect/imagination for the less close friends. Most of these negative coefficients increase (less negative) monotonically with closeness. We also find cases in which respondents are more similar to their least close friends. That is the case of behavioural problems at 12-13 years old, locus of control and novelty seeking.

Estimates by the length of the friendship indicate that recent friends tend to be less similar to respondents, compared to whole-life friends (see for example, the negative and significant coefficients of KS2, IQ, or playing a musical instrument). This is especially true for personality measures (extraversion, emotional stability, and intellect/imagination). These results can be explained by the fact that we do not observe broken links (or friendships). Therefore, it is reasonable to think that very long friendships (longer than 13 years) at this age (15-17 years old) are very solid, and similarities (homophily) in personality between these friends should be stronger. Furthermore, these friends could have influenced each other and shaped their personality together.

The inclusion of length and closeness variables has diverse effects on the average degree of alikeness of friends (β). By adding more (relevant) variables to the model, we reduce the risk of incurring in omitted variable bias. Therefore, the estimates would represent the relationship between each friend characteristic and the characteristic of the respondent, independently of how close they are and how long the friendship has been. Most coefficients remain of the same size and significance. However, their magnitude increase for social skills measures[38], fat mass, fat percentage and bad teen behaviour (arriving late to classes) and become (weakly) significant for some of them. This indicates that we are capturing a higher level of homophily (similarity) in these characteristics when we control for closeness or length of friendship (Tables 11 and 13).

After controlling for common school characteristics (fixed effects), we find very similar results for the length and closeness of friendships. Homophily is stronger among close friends in academic achievement, total IQ, playing a musical instrument, personality, bad teen be-

---

[37]The models are estimated for existing friendships only (i.e. children who were already friends at the time of the measure).

[38]Child makes friends easily, has lots of friends and has more friends than most other kids.

haviour, likelihood of going to university and intensity seeking (Table 14). Respondents are less similar to their closer friends in terms of behavioural problems at 12, locus of control and novelty seeking. The older the friendship, the higher the alikeness in IQ (at 16), playing a musical instrument, personality (extraversion, agreeableness and intellect/imagination mainly) and intensity seeking (at 14). Recent friends are more similar in terms of behavioral problems (at 7 and 13), fat percentage and novelty seeking (at 17). Similarities in academic achievement show an U-shaped homophilic "pattern" in which the scores of the newest and oldest friends are more similar. This result is a strong sign of a deliberate selection process of friends in adolescence.

The degree of alikeness of friends remain, again, very similar after adding length and closeness to the model. Only the coefficients of popularity, playing a musical instrument and arriving late to lessons notably increase and/or become significant (Tables 12 and 14).

5.5 Individual Fixed Effects

We estimate equation (4) for those outcomes that were measured more than once, i.e. KS scores, IQ, locus of control, intensity seeking, body fat percentage and total difficulties score. Table 15 contains the IFE results. After controlling for all unobservables at the individual level, we find a positive and significant degree of alikeness in KS scores and intensity seeking (personality). This might indicate that adolescents consider academic achievement and similar tastes very important factors for the selection of their friends.

In summary, our results indicate that adolescents are very similar to their friends in their academic achievement (especially in KS2 and KS3 scores), total IQ, likelihood of going to university, popularity, bad teen behaviour (arriving late to and skipping classes) and parental education. In addition, some dimensions of personality such as extraversion, locus of control and intensity seeking are very important for teenagers to form friendships. Similarities in academic achievement, total IQ, social skills, popularity, personality, bad teen behaviour, likelihood of going to university and intensity seeking are stronger the closer the friend is to the respondent. Likewise, recently-met friends are less similar to respondents in some dimensions (academic achievement, IQ, playing musical instrument and extraversion, emotional stability, intellect/imagination) compared to whole-life friends. As these are very long friendships (longer than 13 years) this might indicate mutual influence along time.

6 Conclusions

This study analyses the characteristics of a friendship network of adolescents. It is based on the common observation that people tend to establish relationships with other people

who share their similar characteristics or attributes (homophily). We use a new and unique longitudinal dataset that contains education, health, development and family information of a friendship network of 6,961 links.

We estimate a range of specifications to analyse the degree of alikeness of friends, in terms of their academic achievement, ability, personality, behaviour, aspirations and socioeconomic status of their families. We control for unobserved characteristics by using school or individual fixed effects and explore the role of length and closeness of friendships on the degree of homophily. We also exploit the dynamics of the friendship by comparing similarities among existing and future friends.

Characterising homophily is very important for understanding the type, composition and structure of the network (and the society) we are interested in. Our findings show that our network is more structured along ability levels (academic achievement and IQ) than socio-economic status grounds. This last result is surprising as previous research has found strong socio-economic ties (Bearman et al.; 2004; Mesch and Talmud; 2007). Also, personality is a key determinant for the friendships formation process, especially among non-school friends, while physical characteristics such as body mass or body fat are not.

We deal with the problem of different pools of potential friends in two ways, controlling for school fixed effects and drawing a random sample of friends for each respondent (simulated friends), among a pool of potential friends. We show that the degree of homophily among these simulated friends, is much lower than that of the observed friendships. Therefore, we are confident that our results arise from choices rather than differences in the pools of available friends.

Our findings are pertinent for understanding the role of friendships in adolescent society. It is undeniable that people select and influence each other; thus, social networks are powerful in spreading information, beliefs and behaviors. An immediate consequence of homophily is segregation (Centola et al.; 2007; Golub and Jackson; 2011). In our network, high levels of segregation in key characteristics (academic achievement, bad behaviour and educational aspirations) might have future consequences for education attainment, labour opportunities or social mobility. We hope to explore implications for education and social mobility in further work.

Additionally, high levels of homophily slow down the formation of a broad consensus on issues, while promoting high levels of consensus in tight-knit but isolated groups. This is so, as homophily affects the transmission and diffusion of information across –but not within– groups (Golub and Jackson; 2009, 2010). This effect might have important consequences for the promotion of positive behaviours, attitudes and aspirations among adolescents, and also might limit the spread of risk behaviours. It might also have implications for public policy as the structure of the network might influence its expected impact.

References

Altermatt, E. and Pomerantz, E. (2003). The development of competence-related and motivational beliefs: An investigation of similarity and influence among friends, *Journal of Educational Psychology* **95**(1): 111–123.

Baerveldt, C., Van Duijn, M., Vermeij, L. and Van Hemert, D. (2004). Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks, *Social Networks* **26**(1): 55–74.

Bearman, P. and Moody, J. (2004). Suicide and friendships among American adolescents, *American Journal of Public Health* **94**(1): 89.

Bearman, P., Moody, J. and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks, *The American Journal of Sociology* pp. 44–91.

Calvo-Armengol, A. and Jackson, M. (2004). The effects of social networks on employment and inequality, *The American Economic Review* **94**(3): 426–454.

Calvo-Armengol, A. and Jackson, M. (2009). Like father, like son: Social network externalities and parent-child correlation in behavior, *American Economic Journal: Microeconomics* **1**(1): 124–150.

Centola, D., González-Avella, J., Eguíluz, V. and San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups, *Journal of Conflict Resolution* **51**(6): 905.

Cohen, J. (1977). Sources of peer group homogeneity, *Sociology of Education* **50**(4): 227–241.

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

Currarini, S. and Vega-Redondo, F. (2010). Homophily and Search.
URL: *http://venus.unive.it/currarin/home.html*

Doyle, J. and Kao, G. (2007). Friendship choices of multiracial adolescents: Racial homophily, blending, or amalgamation?, *Social science research* **36**(2): 633–653.

Fong, E. and Isajiw, W. (2000). Determinants of friendship choices in multiethnic society, *Sociological Forum* **15**(2): 249–271.

Foster, G. (2005). Making friends: A nonexperimental analysis of social pair formation, *Human Relations* **58**(11): 1443.

Fowler, J. and Christakis, N. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study, *British Medical Journal* **337**(Dec).

French, D., Jansen, E., Riansari, M. and Setiono, K. (2003). Friendships of Indonesian children: Adjustment of children who differ in friendship presence and similarity between mutual friends, *Social Development* **12**(4): 605–621.

Galeotti, A., Ghiglino, C. and Squintani, F. (2009). Strategic information transmission in networks, *Preprint* .

Goeree, J., McConnell, M., Mitchell, T., Tromp, T. and Yariv, L. (2010). The 1/d law of giving, *American Economic Journal: Microeconomics* **2**(1): 183–203.

Golub, B. and Jackson, M. (2009). How homophily affects learning and diffusion in networks, Working Paper.
URL: *http://arxiv.org/pdf/0811.4013*

Golub, B. and Jackson, M. (2010). How Homophily Affects the Speed of Contagion, Best Response and Learning Dynamics, *Working paper* .

Golub, B. and Jackson, M. (2011). Network Structure and the Speed of Learning: Measuring Homophily Based on its Consequences, *Annals of Economics and Statistics, forthcoming, available at http://stanford. edu/~ bgolub/papers/DWH. pdf* .

Gregg, P., Propper, C. and Washbrook, E. (2007). *Understanding the relationship between parental income and multiple child outcomes: a decomposition analysis*, Centre for Analysis of Social Exclusion, London School of Economics and Political Science.

Hamm, J. (2000). Do Birds of a Feather Flock Together? The Variable Bases for African American, Asian American, and European American Adolescents' Selection of Similar Friends?, *Developmental Psychology* **36**(2): 209–219.

Haynie, D. (2002). Friendship networks and delinquency: The relative nature of peer delinquency, *Journal of Quantitative Criminology* **18**(2): 99–134.

Hitsch, G., Hortacsu, A. and Ariely, D. (2010). Matching and sorting in online dating, *The American Economic Review* **100**(1): 130–163.

Jackson, M. (2007). Social Structure, Segregation, and Economic Behavior, *Presented as the Nancy Schwartz Memorial Lecture in April of 2007* .

Jackson, M. (2008). *Social and economic networks*, Princeton University Press.

Jackson, M. (2010). An Overview of Social Networks and Economic Applications, *in Handbook of Social Economics, edited by J. Benhabib, A. Bisin and M. Jackson* .

Kandel, D. (1978). Homophily, selection, and socialization in adolescent friendships, *The American Journal of Sociology* **84**(2): 427.

Kao, G. and Joyner, K. (2004). Do race and ethnicity matter among friends?, *Sociological quarterly* **45**(3): 557–573.

Kossinets, G. and Watts, D. (2009). Origins of Homophily in an Evolving Social Network, *American Journal of Sociology* **115**(2): 405–50.

Lazarsfeld, P. and Merton, R. (1954). Friendship as a social process: A substantive and methodological analysis, *Freedom and control in modern society* **18**: 66.

Lundborg, P. (2006). Having the wrong friends? Peer effects in adolescent substance use, *Journal of Health Economics* **25**(2): 214–233.

Marmaros, D. and Sacerdote, B. (2006). How do Friendships Form?*, *Quarterly Journal of Economics* **121**(1): 79–119.

Marsden, P. (1988). Homogeneity in confiding relations, *Social networks* **10**(1): 57–76.

McPherson, J. and Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups, *American Sociological Review* **52**: 370–379.

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

Mesch, G. and Talmud, I. (2007). Similarity and the quality of online and offline social relationships among adolescents in Israel, *Journal of Research on Adolescence* **17**(2): 455–465.

Mollica, K., Gray, B. and Trevino, L. (2003). Racial homophily and its persistence in newcomers' social networks, *Organization Science* **14**(2): 123–136.

Moody, J. (2001). Race, School Integration, and Friendship Segregation in America, *The American Journal of Sociology* **107**(3): 679–716.

Pearson, M., Steglich, C. and Snijders, T. (2006). Homophily and assimilation among sport-active adolescent substance users, *Connections* **27**(1): 51–67.

Prinstein, M., Boergers, J. and Spirito, A. (2001). Adolescents' and their friends' health-risk behavior: factors that alter or add to peer influence, *Journal of pediatric psychology* **26**(5): 287.

Rogers, E. and Bhowmik, D. (1970). Homophily-heterophily: Relational concepts for communication research, *Public Opinion Quarterly* **34**(4): 523–538.

Rostila, M. (2010). Birds of a Feather Flock Together–and fall ill? Migrant homophily and health in Sweden, *Sociology of Health and Illness* **32**(3): 382–399.

Shrum, W., Cheek Jr, N. and MacD, S. (1988). Friendship in school: Gender and racial homophily, *Sociology of Education* **61**(4): 227–239.

Skopek, J., Schulz, F. and Blossfeld, H. (2010). Who Contacts Whom? Educational Homophily in Online Mate Selection, *European Sociological Review* pp. 1–16.

The ALSPAC Study Team (2010). The ALSPAC Study. Data collected at Teen Focus 2 (around 13 $\frac{1}{2}$ years). Documentation giving frequencies, background and instructions for use., *Technical report*, University of Bristol. February. Last updated for version 1b of the release file.

Wimmer, A. and Lewis, K. (2010). Beyond and below racial homophily. ERG models of a friendship network documented on Facebook, *Forthcoming in American Journal of Sociology* **116**.

Figures

Figure 1: Distribution of closeness measure by nominated friend.
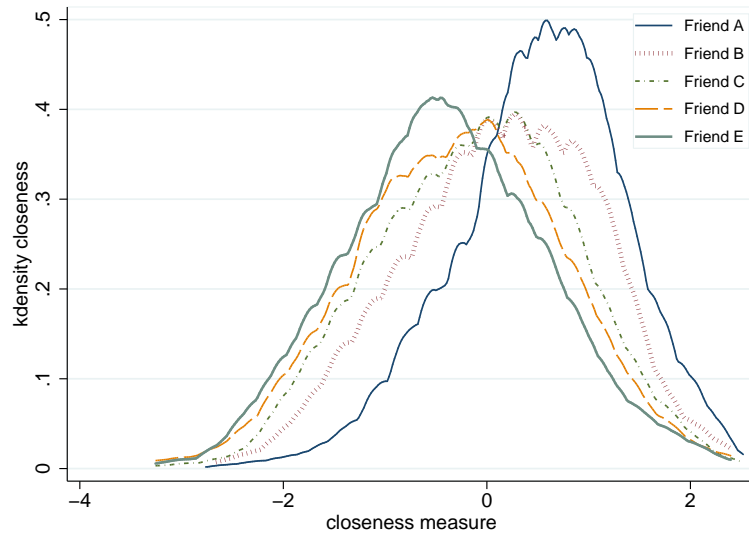
Figure 2: Homophily among respondents, friends and simulated friends. Academic achievement

(a) Real friends.

(b) Simulated friends.
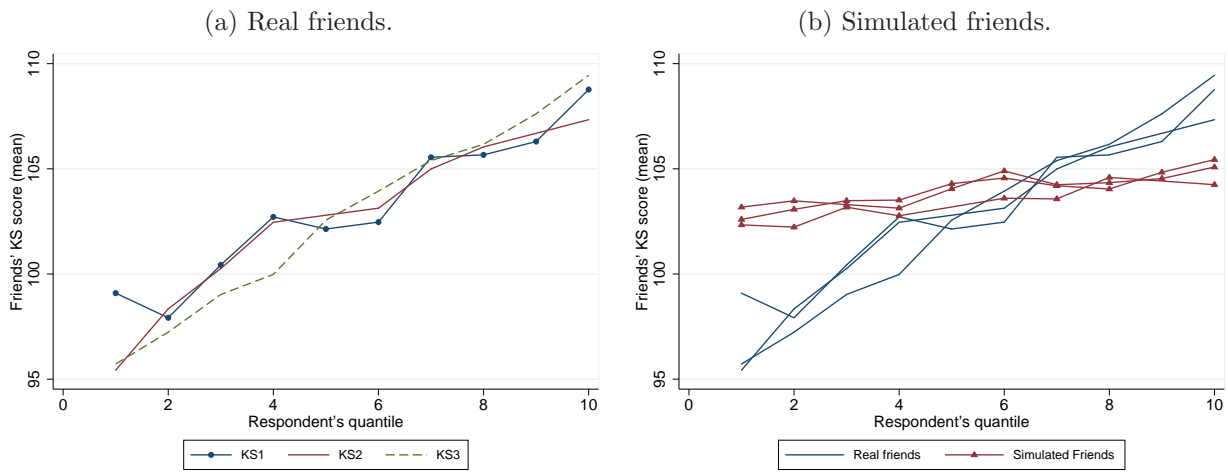
Figure 3: Homophily among respondents, friends and simulated friends. IQ

(a) Real friends.

(b) Simulated friends.
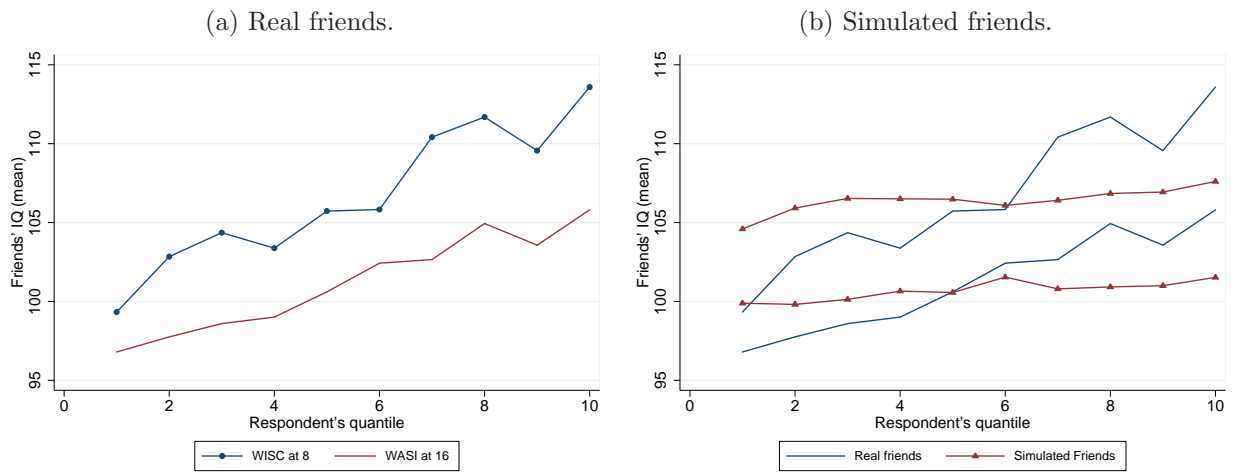
Figure 4: Homophily among respondents, friends and simulated friends. Personality

(a) Real friends.

(b) Simulated friends.
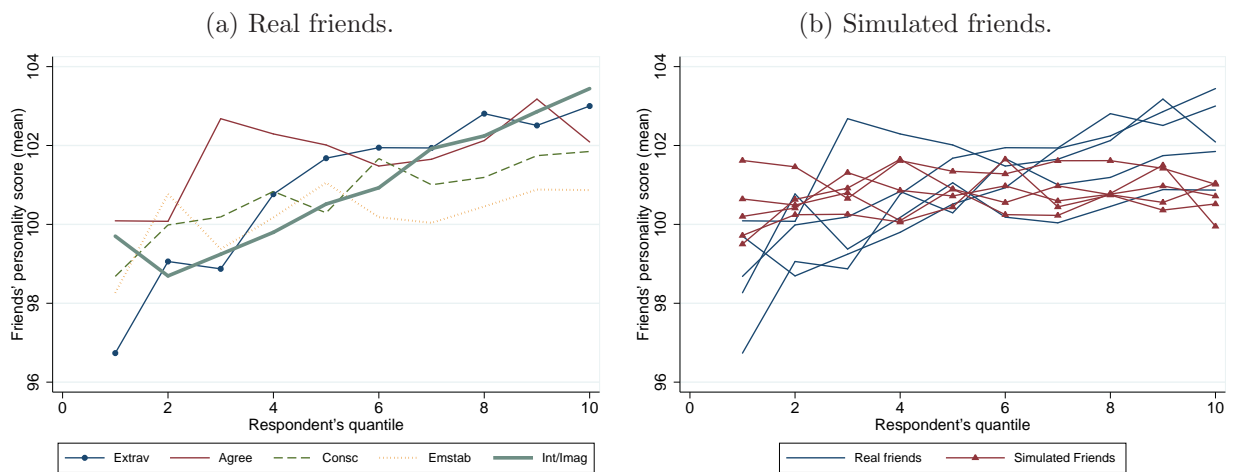
Figure 5: Homophily among respondents, friends and simulated friends. Intensity seeking

(a) Real friends.

(b) Simulated friends.

27
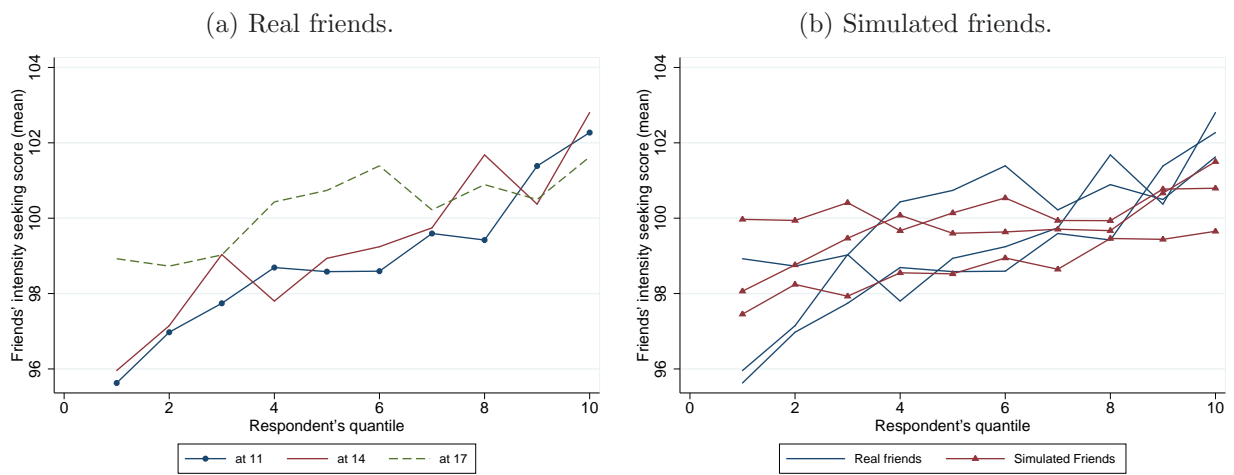
Figure 6: Homophily among respondents, friends and simulated friends. Locus of control

(a) Real friends.

(b) Simulated friends.
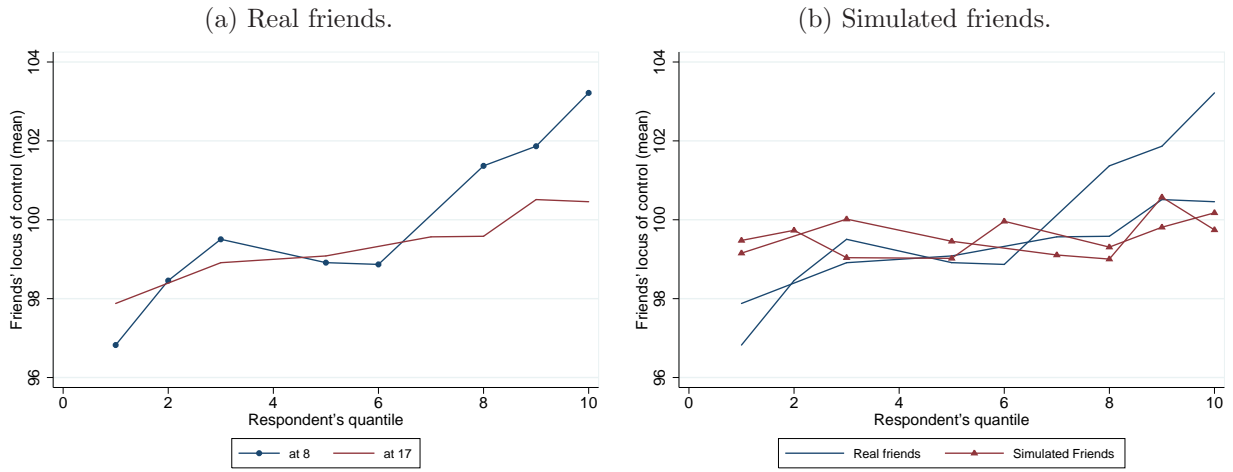
Figure 7: Homophily among respondents, friends and simulated friends. Body fat

(a) Real friends.

(b) Simulated friends.
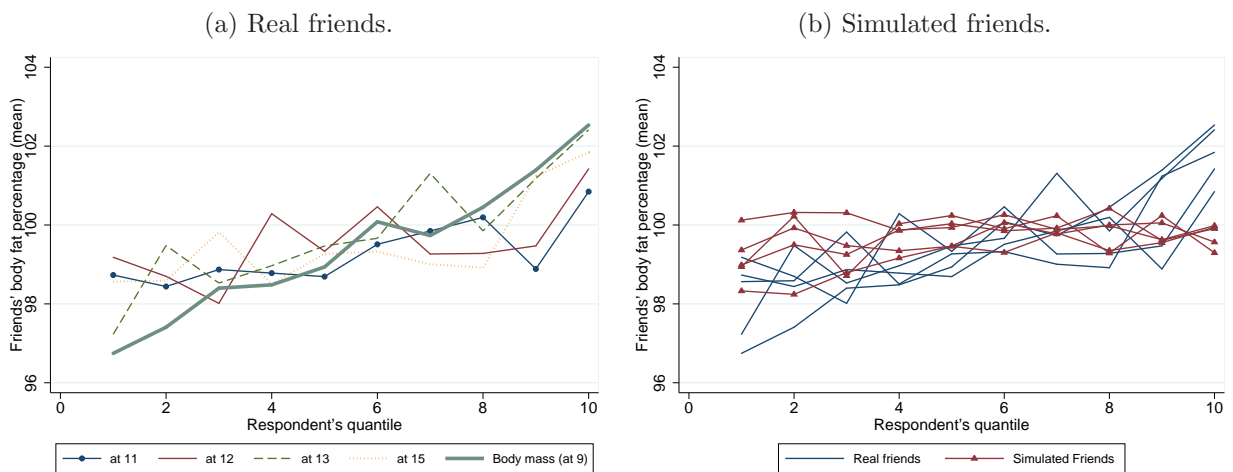
Tables

Table 1: Frequency of nominations - ALSPAC network

No. of nominated friends	No. of respondends	%	Total links	%
1	484	20.2	484	7.0
2	496	20.7	992	14.3
3	522	21.8	1566	22.5
4	551	23.0	2204	31.7
5	343	14.3	1715	24.6
Total	**2,396**	**100.0**	**6,961**	**100.0**

Table 2: Total scores

groups	Total score				
(g)	Obs	Mean	Std. Dev.	Min	Max
boy-boy same school	2,919	10.8	3.8	1	20
boy-boy diff school	487	10.0	3.5	1	18
girl-girl same school	5,784	11.9	3.4	0	20
girl-girl diff school	971	10.3	3.1	1	18
opposite sex same school	1,973	10.4	3.8	1	20
opposite sex diff school	895	10.3	3.5	1	18
Total observations	13,029				

Table 3: Closeness measure. Descriptive statistics.

Friend	closeness				
	Obs	Mean	Std. Dev.	Min	Max
A	2,762	0.52	0.84	-2.89	2.52
B	2,748	0.13	0.96	-3.18	2.43
C	2,693	-0.12	0.98	-3.18	2.52
D	2,519	-0.27	0.99	-3.48	2.43
E	2,307	-0.34	0.98	-3.18	2.50

Table 4: Respondents and non-respondents comparison

	Completed questionnaire		Total	Two-sample t-test with equal variances
	Yes	**No**		
	(1)	(2)	(3)	(4)
gender				
male	36.62	57.38	49.12	
female	63.38	42.62	50.88	18.41*
ethnicity				
white	96.99	94.69	95.65	
non-white	3.01	5.31	4.35	-4.64*
Anthropometric				
Height (cms) (15 years old)	168.3	170.0	169.1	-6.48*
Fat mass[1] (9 years old)	100.5	99.9	100.2	2.27*
Obese				
No	89.7	89.7	89.7	
Yes	10.3	10.3	10.3	0.080
IQ and school performance				
KS1 score[2]	16.6	15.2	15.7	16.8*
KS2 score[3]	29.3	27.7	28.3	17.6*
KS3 score[4]	38.5	35.1	36.4	20.3*
Total IQ (8 years old)	106.9	102.0	104.4	10.9*
Behaviour				
Total difficulties score (4 years old)	98.5	100.1	99.4	-6.6*
Total difficulties score (7 years old)	98.4	100.6	99.6	-8.05*
Total difficulties score (12 years old)	98.5	100.8	99.7	-8.66*
Total difficulties score (13 years old)	98.7	100.7	99.7	-6.76*
self esteem				
Total self esteem score (8 years old)	18.2	18.1	18.2	1.69
Personality				
Extraversion[5]	35.2	35.5	35.3	-1.58
Agreeableness[5]	35.6	34.8	35.2	6.32*
Conscientiousness[5]	32.3	31.5	31.9	4.66*
Emotional stability[5]	31.7	31.4	31.6	1.48
Openness[5]	36.2	35.6	35.9	3.19*
Locus of control (17 years old)	99.8	100.5	100.0	-2.15*
Locus of control (8 years old)	99.4	100.3	99.9	-2.91*
Intensity[6]	25.1	25.1	25.1	0.09
Novelty[6]	23.0	23.3	23.1	-2.19*

[1] adjusted for age, height and height2 (std).

[2] average for reading, writing, spelling and maths.

[3] average for reading, writing, science and maths.

[4] average of attainment point scores for English, Maths and Science.

[5] Five Factor Model (FFM) see Section 3.3

[6] Arnetts Inventory for Sensation Seeking see Section 3.3

Table 5: Respondents and non-respondents comparison

	Completed questionnaire		Total	Two-sample t-test with equal variances
	Yes	No		
	(1)	(2)	(3)	(4)
Friends score (8 years old) 15=good, 0=bad	11.50	11.60	11.55	-1.45
Friends score (10 years old) 17=good, 0=bad	14.00	14.04	14.02	-0.77
Number of close friends (10 years old)	4.81	4.99	4.90	-1.87
Number of close friends (11 years old)	5.34	5.66	5.50	-3.01*
Number of close friends (13 years old)	5.30	5.50	5.39	-1.97*
Makes friends easily (9 years old)	72.51	73.44	72.99	-0.79
Has lots of friends (9 years old)	86.12	86.91	86.53	-0.86
Child has more friends than most other kids (9 years old)	37.89	44.39	41.25	-4.93*
Child talks about school friends (10-11 years old)	83.99	83.29	83.63	0.70
Child talks about school friends (13-14 years old)	77.91	73.59	75.8	3.6*
Child is rather solitary (11-12 years old)	23.5	25.78	24.65	-1.93
Child is rather solitary (13 years old)	29.58	30.65	30.11	-0.84

The definitions of these variables are in Section 3.3

Table 6: Respondents and non-respondents comparison

| | Completed questionnaire | | Total | Two-sample t-test with equal variances |
| | Yes | No | | |
	(1)	(2)	(3)	(4)
Mother's education				
CSE/none	11.39	20.82	16.98	-10.67*
voc/O level	41.37	45.74	43.96	-3.71*
A level	27.65	22.47	24.58	5.08*
degree	19.59	10.97	14.48	10.4*
Father's education				
CSE/none	16.72	25.66	21.98	-9.05*
voc/O level	28.2	31.48	30.13	-2.98*
A level	29.45	26.99	28.01	2.28*
degree	25.63	15.86	19.89	10.27*
Mother's occupational class				
Professional	8.29	5.34	6.61	4.55*
Managerial	36.31	30.44	32.95	4.78*
Skilled non-manual	41	43.76	42.58	-2.13*
Skilled manual	6.22	8.07	7.28	-2.72*
Semi-skilled manual	6.85	10.24	8.79	-4.59*
Non-skilled manual	1.33	2.14	1.79	-2.34*
Father's occupational class				
Professional	15.06	10.26	12.3	5.81*
Managerial	38.16	33.54	35.51	3.83*
Skilled non-manual	11.95	11.15	11.49	0.99
Skilled manual	26.24	32.05	29.57	-5.05*
Semi-skilled manual	7.35	10.07	8.91	-3.79*
Non-skilled manual	1.24	2.93	2.21	-4.56*
House ownership (at 8/21/33/61 months)				
Owned throughout 8-61 months	67.31	47.84	55.89	16.73*
Ever in other	22.64	32.08	28.18	-8.83*
Ever in social housing	10.05	20.08	15.94	-11.58*
Income				
financial difficulties score 8/21/33/61 months	2.50	3.25	2.94	-10.3*
ln(average equivalised income[1])	5.40	5.26	5.32	11.3*
Income quintiles				
1	16.02	23.94	20.55	-7.79*
2	17.81	21.04	19.66	-3.22*
3	20.62	20.96	20.81	-0.33
4	22.08	18.14	19.83	3.92*
5	23.47	15.92	19.16	7.62*

[1] The definitions of the variables are in Section 3.3

Table 7: Homophily estimations. Five real and simulated friends. OLS results.

Variable	Mean of var		N	Real Friend					Mean of var**		N	Simulated friend**				
	Resp	Friends		A	B	C	D	E	Resp	Friends		A	B	C	D	E
				(1)	(2)	(3)	(4)	(5)				(6)	(7)	(8)	(9)	(10)
Mother's education	2.5	2.4	4,231	0.283*	0.283*	0.284*	0.279*	0.282*	2.4	2.4	7,442	0.181*	0.182*	0.183*	0.182*	0.071
				[0.033]	[0.033]	[0.034]	[0.034]	[0.034]				[0.022]	[0.022]	[0.022]	[0.022]	[0.081]
Father's education	2.6	2.5	4,101	0.238*	0.238*	0.238*	0.241*	0.230*	2.5	2.5	7,273	0.154*	0.158*	0.156*	0.159*	0.171†
				[0.034]	[0.035]	[0.035]	[0.036]	[0.036]				[0.023]	[0.023]	[0.023]	[0.023]	[0.095]
ln(equivalised income)	5.4	5.4	3,269	0.182*	0.178*	0.179*	0.181*	0.179*	5.4	5.4	6,393	0.129*	0.129*	0.128*	0.128*	0.137*
(at 33 and 47 months)				[0.030]	[0.030]	[0.030]	[0.030]	[0.030]				[0.02]	[0.02]	[0.02]	[0.02]	[0.026]
Financial difficulties score	2.4	2.6	4,057	0.039	0.059	-0.014	0.05	0.027	2.6	2.7	7,405	0.05	0.055	0.052	0.053	-0.001
(8/21/33/61 months)				[0.056]	[0.055]	[0.054]	[0.057]	[0.065]				[0.027]	[0.027]	[0.027]	[0.027]	[0.121]
Mothers's occup. class	2.7	2.8	3,016	0.150*	0.144*	0.134*	0.141*	0.126*	2.8	2.8	5,917	0.093*	0.093*	0.091*	0.091*	0.135†
(at baby's birth)				[0.039]	[0.040]	[0.040]	[0.042]	[0.041]				[0.024]	[0.024]	[0.024]	[0.024]	[0.064]
Fathers's occup. class	2.8	2.8	3,497	0.156*	0.130*	0.151*	0.153*	0.136*	2.9	2.9	6,467	0.114*	0.112*	0.112*	0.112*	0.209
(at baby's birth)				[0.035]	[0.036]	[0.036]	[0.037]	[0.038]				[0.024]	[0.024]	[0.024]	[0.024]	[0.121]
KS3 score	106.1	105.0	3,929	0.383*	0.382*	0.384*	0.380*	0.384*	105.6	104.6	7,013	0.092*	0.092*	0.092*	0.092*	0.131*
				[0.033]	[0.033]	[0.033]	[0.033]	[0.033]				[0.022]	[0.022]	[0.022]	[0.022]	[0.037]
Extraversion	100.0	100.7	2,176	0.207*	0.204*	0.209*	0.206*	0.206*	99.9	100.3	5,173	-0.012	-0.012	-0.011	-0.012	0.005
				[0.044]	[0.044]	[0.044]	[0.044]	[0.044]				[0.025]	[0.025]	[0.025]	[0.025]	[0.034]
Agreeableness	101.9	101.9	2,094	-0.024	-0.026	-0.03	-0.029	-0.027	101.3	101.1	5,103	-0.018	-0.018	-0.018	-0.018	-0.019
				[0.044]	[0.044]	[0.044]	[0.045]	[0.045]				[0.024]	[0.024]	[0.024]	[0.024]	[0.034]
Conscientiousnes	101.6	101.0	2,000	0.013	0.015	0.014	0.015	0.012	101.2	100.6	4,920	-0.003	-0.003	-0.003	-0.003	-0.017
				[0.039]	[0.039]	[0.039]	[0.039]	[0.039]				[0.024]	[0.024]	[0.024]	[0.024]	[0.035]
Emotional Stability	100.2	100.4	2,037	-0.016	-0.026	-0.021	-0.018	-0.022	100.4	100.2	4,950	-0.02	-0.02	-0.02	-0.02	-0.022
				[0.038]	[0.038]	[0.038]	[0.038]	[0.038]				[0.023]	[0.023]	[0.023]	[0.023]	[0.032]
Intellect/Imagination	101.0	101.0	2,093	0.092*	0.084*	0.087*	0.084*	0.087*	100.9	100.6	5,073	0.004	0.003	0.003	0.004	0.012
				[0.039]	[0.039]	[0.039]	[0.039]	[0.039]				[0.024]	[0.024]	[0.024]	[0.024]	[0.033]
Locus of control	98.8	99.0	2,096	0.089†	0.088†	0.096*	0.093*	0.094*	99.6	99.7	4,529	0.012	0.012	0.012	0.012	0.015
				[0.046]	[0.046]	[0.045]	[0.046]	[0.046]				[0.027]	[0.027]	[0.027]	[0.027]	[0.032]
Intensity seeking	25.3	25.3	2,189	0.091†	0.078	0.097*	0.091†	0.080†	25.3	25.2	4,853	-0.012	-0.011	-0.011	-0.012	0.014
				[0.048]	[0.049]	[0.048]	[0.048]	[0.048]				[0.027]	[0.027]	[0.027]	[0.027]	[0.038]
Novelty seeking	23.1	22.9	2,147	0.131*	0.135*	0.140*	0.143*	0.135*	23.2	23.1	4,754	-0.003	-0.002	-0.003	-0.002	0.023
				[0.043]	[0.044]	[0.044]	[0.043]	[0.044]				[0.027]	[0.027]	[0.027]	[0.027]	[0.043]

* Significant at 5%, †Significant at 10%. Robust standard errors in brackets.
** Means of 130 simulations which randomly selected five friends from a sample of people (real friends and respondent not included) who attended the same school and academic year as the respondent. Four of them are of the same gender and one of the opposite gender. Friend E is always the opposite gender friend.

Table 8: Homophily estimations. Five real and simulated friends. OLS results. F-test and other covariates.

Variable	F-test all β_j's equal F / p-value	Real Friend Female respondent (2)	Gender composition GG (3)	GB (4)	BG (5)	Resp. age (6)	Friend's age (7)	Same school (8)	F-test all β_j equal F / p-value	Simulated Friend** Female respondent (10)	Gender composition GG (11)	GB (12)	BG (13)	Resp. age (14)	Friend's age (15)
Mother's education	0.03	-0.094	0.038	0.012	-0.008	-0.006†	0.006*	-0.060*	0.80	0.067	-0.061†	0.048	0.112	-0.004	0.002
	0.998	[0.088]	[0.035]	[0.038]	[0.022]	[0.003]	[0.003]	[0.014]	0.572	[0.062]	[0.025]	[0.083]	[0.077]	[0.003]	[0.002]
Father's education	0.09	-0.188*	0.061†	0.032	-0.033	-0.007†	0.002	-0.039*	0.19	0.04	-0.035	-0.056	-0.008	-0.006*	0.004
	0.986	[0.092]	[0.035]	[0.039]	[0.025]	[0.004]	[0.004]	[0.016]	0.930	[0.067]	[0.026]	[0.096]	[0.092]	[0.003]	[0.003]
ln(equivalised income) (at 33 and 47 months)	0.39	-0.293	0.051	0.045	-0.012*	-0.005*	0.005*	-0.006	0.20	0.153	-0.034	-0.045	-0.01	-0.004*	0.001
	0.819	[0.199]	[0.037]	[0.037]	[0.006]	[0.002]	[0.002]	[0.004]	0.922	[0.138]	[0.026]	[0.031]	[0.017]	[0.001]	[0.001]
Financial difficulties score (8/21/33/61 months)	1.26	-0.265*	0.086*	0.118*	0.009	0.040*	-0.011	0.01	0.58	0.149	-0.01	0.037	0.047	0.052*	-0.008
	0.283	[0.130]	[0.041]	[0.052]	[0.049]	[0.011]	[0.010]	[0.038]	0.684	[0.093]	[0.026]	[0.123]	[0.118]	[0.009]	[0.008]
Mothers's occup. class (at baby's birth)	0.41	0.05	0.005	0.022	-0.049*	0.000	-0.003	-0.004	0.51	0.241*	-0.035	-0.069	-0.046	0.003	-0.001
	0.800	[0.118]	[0.042]	[0.045]	[0.024]	[0.005]	[0.004]	[0.019]	0.731	[0.081]	[0.028]	[0.067]	[0.059]	[0.003]	[0.003]
Fathers's occup. class (at baby's birth)	0.66	0.01	0.046	0.067†	0.049†	0.011*	-0.010*	0.029	0.42	0.014	0.032	-0.063	-0.098	0.002	-0.002
	0.620	[0.108]	[0.036]	[0.040]	[0.028]	[0.005]	[0.005]	[0.019]	0.785	[0.082]	[0.027]	[0.122]	[0.118]	[0.004]	[0.004]
KS3 score	0.48	-5.487	0.052	0.056	0.010*	0.134*	-0.025	0.011*	0.60	1.026	-0.009	-0.048	-0.039	0.116*	-0.066*
	0.751	[3.882]	[0.037]	[0.037]	[0.005]	[0.028]	[0.027]	[0.004]	0.669	[2.717]	[0.026]	[0.039]	[0.03]	[0.023]	[0.023]
Extraversion	0.17	4.703	-0.023	-0.015	0.022*	0.068	0.019	-0.002	0.27	0.145	0.025	0.005	-0.016	0.08*	0.033
	0.953	[5.086]	[0.051]	[0.051]	[0.009]	[0.054]	[0.051]	[0.006]	0.886	[3.002]	[0.03]	[0.038]	[0.023]	[0.035]	[0.035]
Agreeableness	0.32	2.862	0.023	0.031	0.017†	0.036	-0.02	-0.005	0.12	3.872	0.014	0.014	0.001	-0.041	0.001
	0.867	[5.017]	[0.050]	[0.051]	[0.010]	[0.050]	[0.048]	[0.005]	0.967	[2.946]	[0.029]	[0.038]	[0.023]	[0.033]	[0.032]
Conscientiousnes	0.10	-13.921*	0.122*	0.127*	0.005	0.019	-0.085	0.002	0.21	-2.1	0.004	0.02	0.013	-0.001	0.027
	0.983	[4.668]	[0.046]	[0.046]	[0.009]	[0.056]	[0.054]	[0.007]	0.916	[2.995]	[0.03]	[0.039]	[0.026]	[0.036]	[0.035]
Emotional Stability	0.87	-12.597*	0.076	0.076	-0.009	0.144*	-0.008	0.002	0.08	-8.326*	0.033	0.034	0.002	0.103*	-0.041
	0.483	[4.758]	[0.047]	[0.047]	[0.009]	[0.053]	[0.051]	[0.006]	0.984	[3.001]	[0.03]	[0.036]	[0.022]	[0.036]	[0.035]
Intellect/Imagination	0.52	-9.841*	0.086†	0.091†	0.015†	0.089	0.018	0.003	0.12	0.063	-0.013	-0.022	-0.009	0.081*	0.015
	0.721	[4.760]	[0.047]	[0.047]	[0.008]	[0.056]	[0.054]	[0.007]	0.969	[3.029]	[0.03]	[0.038]	[0.023]	[0.036]	[0.035]
Locus of control	0.56	-0.083	0.012	0.007	0.009	-0.06	-0.005	-0.011†	0.08	1.786	-0.005	-0.007	-0.003	-0.092*	-0.02
	0.693	[5.104]	[0.052]	[0.053]	[0.009]	[0.051]	[0.050]	[0.006]	0.986	[3.25]	[0.033]	[0.037]	[0.016]	[0.039]	[0.038]
Intensity seeking	1.42	3.093*	-0.051	-0.069	-0.014	0.012	-0.004	0.006	0.31	1.394	0.015	-0.009	-0.023	0.006	0.007
	0.224	[1.331]	[0.054]	[0.056]	[0.013]	[0.018]	[0.018]	[0.008]	0.864	[0.78]	[0.031]	[0.042]	[0.027]	[0.013]	[0.013]
Novelty seeking	0.34	1.894†	-0.062	-0.082	-0.047*	-0.041*	-0.012	-0.012	0.26	0.692	-0.003	-0.026	-0.023	-0.031*	0.01
	0.851	[1.120]	[0.049]	[0.051]	[0.014]	[0.020]	[0.020]	[0.010]	0.898	[0.739]	[0.032]	[0.046]	[0.033]	[0.015]	[0.015]

* Significant at 5%. †Significant at 10%. Robust standard errors in brackets.
** Means of 130 simulations which randomly selected five friends from a sample of people (real friends and respondent not included) who attended the same school and academic year as the respondent. Four of them are of the same gender and one of the opposite gender. Friend E is always the opposite gender friend.

34

Table 9: Homophily estimations. Five real and simulated friends. School fixed effects results.

Variable	Mean of var		N	Real Friend					Mean of var**		N	Simulated friend**				
				A	**B**	**C**	**D**	**E**				**A**	**B**	**C**	**D**	**E**
	Resp	Friends	N_{Sch}	(1)	(2)	(3)	(4)	(5)	Resp	Friends	N_{Sch}	(6)	(7)	(8)	(9)	(10)
Mother's education	2.5	2.4	3,950	0.108*	0.102*	0.114*	0.101*	0.113*	2.4	2.4	7,401	-0.016	-0.016	-0.014	-0.015	-0.123
			63	[0.035]	[0.035]	[0.034]	[0.036]	[0.036]			55	[0.022]	[0.022]	[0.021]	[0.022]	[0.075]
Father's education	2.6	2.5	3,830	0.086†	0.075	0.091†	0.090†	0.082	2.5	2.5	7,232	-0.014	-0.013	-0.013	-0.011	0.021
			64	[0.051]	[0.049]	[0.053]	[0.049]	[0.054]			55	[0.022]	[0.022]	[0.022]	[0.022]	[0.098]
ln(equivalised income)	5.4	5.4	3,056	0.041	0.037	0.038	0.039	0.038	5.4	5.4	6,359	-0.023	-0.023	-0.023	-0.024	-0.021
(at 33 and 47 months)			63	[0.037]	[0.036]	[0.036]	[0.037]	[0.036]			54	[0.021]	[0.021]	[0.021]	[0.021]	[0.026]
Financial difficulties score	2.5	2.7	3,785	-0.009	0.013	-0.056	-0.004	-0.009	2.6	2.7	7,368	-0.002	0.004	-0.001	0.000	-0.014
(8/21/33/61 months)			64	[0.075]	[0.066]	[0.065]	[0.071]	[0.079]			55	[0.024]	[0.024]	[0.024]	[0.024]	[0.106]
Mothers's occup. class	2.8	2.8	2,793	0.073	0.083	0.057	0.065	0.06	2.8	2.8	5,887	-0.005	-0.005	-0.007	-0.006	0.086
(at baby's birth)			60	[0.052]	[0.052]	[0.053]	[0.054]	[0.053]			55	[0.022]	[0.022]	[0.022]	[0.022]	[0.058]
Fathers's occup. class	2.8	2.9	3,264	0.034	0.005	0.024	0.027	0.009	2.9	2.9	6,434	-0.038	-0.038	-0.037	-0.038	0.025
(at baby's birth)			64	[0.040]	[0.041]	[0.042]	[0.037]	[0.041]			55	[0.023]	[0.023]	[0.023]	[0.023]	[0.12]
KS3 score	106.1	105.0	3,878	0.292*	0.291*	0.292*	0.289*	0.293*	105.6	104.6	7,007	-0.015	-0.015	-0.014	-0.015	-0.014
			55	[0.035]	[0.035]	[0.035]	[0.036]	[0.036]			48	[0.02]	[0.02]	[0.02]	[0.02]	[0.028]
Extraversion	100.0	100.7	2,032	0.175*	0.172*	0.175*	0.174*	0.173*	99.9	100.3	5,160	-0.058*	-0.059*	-0.058*	-0.058*	-0.034
			61	[0.050]	[0.050]	[0.051]	[0.051]	[0.052]			51	[0.024]	[0.024]	[0.024]	[0.024]	[0.037]
Agreeableness	101.9	101.8	1,947	-0.046	-0.047	-0.051	-0.052	-0.046	101.3	101.1	5,087	-0.034	-0.034	-0.034	-0.034	-0.034
			60	[0.057]	[0.055]	[0.056]	[0.057]	[0.057]			52	[0.022]	[0.022]	[0.022]	[0.022]	[0.033]
Conscientiousnes	101.6	101.0	1,859	-0.018	-0.012	-0.013	-0.013	-0.014	101.1	100.6	4,903	-0.051†	-0.051†	-0.051†	-0.051†	-0.046
			60	[0.038]	[0.039]	[0.039]	[0.039]	[0.039]			52	[0.023]	[0.023]	[0.023]	[0.023]	[0.03]
Emotional Stability	100.2	100.3	1,903	-0.045	-0.054	-0.052	-0.049	-0.052	100.4	100.2	4,937	-0.042	-0.042	-0.042	-0.042	-0.028
			61	[0.047]	[0.047]	[0.047]	[0.047]	[0.047]			52	[0.021]	[0.021]	[0.021]	[0.021]	[0.029]
Intellect/Imagination	101.0	100.9	1,949	0.075†	0.069	0.072	0.069	0.07	100.9	100.6	5,056	-0.033	-0.034	-0.034	-0.033	-0.013
			60	[0.045]	[0.045]	[0.046]	[0.045]	[0.047]			52	[0.023]	[0.023]	[0.023]	[0.023]	[0.035]
Locus of control	99.0	99.1	1,928	0.057	0.057	0.065	0.064	0.064	99.6	99.7	4,511	-0.035	-0.035	-0.035	-0.034	-0.022
			60	[0.056]	[0.056]	[0.055]	[0.055]	[0.056]			43	[0.03]	[0.031]	[0.031]	[0.031]	[0.039]
Intensity seeking	25.3	25.3	2,019	0.051	0.042	0.056	0.053	0.047	25.3	25.2	4,826	-0.048	-0.048	-0.048	-0.048	-0.035
			59	[0.049]	[0.052]	[0.051]	[0.049]	[0.051]			45	[0.025]	[0.025]	[0.025]	[0.025]	[0.038]
Novelty seeking	23.1	23.0	1,982	0.102	0.105	0.107	0.114†	0.106†	23.2	23.1	4,729	-0.045	-0.044	-0.045	-0.044	-0.008
			60	[0.061]	[0.063]	[0.066]	[0.062]	[0.063]			45	[0.029]	[0.029]	[0.029]	[0.029]	[0.036]

* Significant at 5%, †Significant at 10%. Robust standard errors in brackets. Clustered at school level.
** Means of 130 simulations which randomly selected five friends from a sample of people (real friends and respondent not included) who attended the same school and academic year as the respondent. Four of them are of the same gender and one of the opposite gender. Friend E is always the opposite gender friend.

Table 10: Homophily estimations. Five real and simulated friends. School fixed effects results. F-test and other covariates.

Variable	F-test all β_j's equal F / p-value	Real Friend Female respondent (2)	GG (3)	GB (4)	BG (5)	Resp. age (6)	Friend's age (7)	Same school (8)	F-test all β_j equal F / p-value	Simulated Friend** Female respondent (10)	GG (11)	GB (12)	BG (13)	Resp. age (14)	Friend's age (15)
Mother's education	0.55	-0.091	0.031	0.011	-0.019	-0.005	0.006†	-0.031†	1.45	-0.04	-0.012	0.096	0.106	-0.005	0.000
	0.701	[0.069]	[0.031]	[0.036]	[0.020]	[0.006]	[0.004]	[0.016]	0.323	[0.069]	[0.025]	[0.075]	[0.073]	[0.006]	[0.003]
Father's education	0.56	-0.165	0.042	0.016	-0.045	-0.004	0.001	-0.025	0.80	0.002	-0.018	-0.053	-0.027	-0.004	0.003
	0.695	[0.101]	[0.051]	[0.051]	[0.028]	[0.006]	[0.003]	[0.017]	0.625	[0.082]	[0.026]	[0.1]	[0.094]	[0.006]	[0.004]
ln(equivalised income)	0.47	-0.337	0.059	0.056	-0.012*	-0.005†	0.004†	-0.001	0.75	0.072	-0.016	-0.019	-0.002	-0.004	0.000
(at 33 and 47 months)	0.756	[0.233]	[0.043]	[0.044]	[0.005]	[0.003]	[0.002]	[0.005]	0.616	[0.145]	[0.027]	[0.032]	[0.018]	[0.003]	[0.002]
Financial difficulties score	1.15	-0.185	0.067	0.078	-0.014	0.040*	-0.006	0.009	1.11	0.207	-0.023	-0.019	0.005	0.049*	-0.004
(8/21/33/61 months)	0.344	[0.184]	[0.053]	[0.059]	[0.054]	[0.016]	[0.010]	[0.044]	0.467	[0.152]	[0.026]	[0.109]	[0.101]	[0.014]	[0.011]
Mothers's occup. class	1.09	0.121	0.003	0.01	-0.057*	0.001	0.001	-0.022	1.98	0.213†	-0.023	-0.114	-0.091	0.004	0.003
(at baby's birth)	0.371	[0.151]	[0.054]	[0.056]	[0.026]	[0.007]	[0.004]	[0.024]	0.186	[0.102]	[0.028]	[0.062]	[0.054]	[0.007]	[0.003]
Fathers's occup. class	1.62	0.009	0.038	0.036	0.029	0.005	-0.008	0.007	0.94	0.032	0.027	-0.04	-0.063	0.001	0.000
(at baby's birth)	0.180	[0.125]	[0.039]	[0.036]	[0.026]	[0.008]	[0.005]	[0.020]	0.528	[0.094]	[0.027]	[0.125]	[0.113]	[0.009]	[0.005]
KS3 score	1.35	-8.435*	0.078†	0.084*	0.008†	0.124*	-0.018	0.013*	0.89	2.043	-0.021	-0.021	-0.001	0.099†	-0.054
	0.265	[4.157]	[0.040]	[0.040]	[0.004]	[0.044]	[0.027]	[0.005]	0.535	[2.665]	[0.025]	[0.032]	[0.018]	[0.049]	[0.032]
Extraversion	0.15	3.531	-0.009	-0.002	0.023*	0.012	0.006	-0.014*	1.23	-0.279	0.029	0.005	-0.022	0.043	-0.012
	0.964	[7.856]	[0.078]	[0.077]	[0.010]	[0.084]	[0.058]	[0.006]	0.452	[2.939]	[0.029]	[0.039]	[0.027]	[0.077]	[0.047]
Agreeableness	1.34	1.031	0.042	0.049	0.015	0.038	-0.046	-0.009	1.10	5.292	0.002	0.000	0.002	-0.044	0.000
	0.266	[5.975]	[0.060]	[0.060]	[0.010]	[0.067]	[0.050]	[0.006]	0.497	[2.773]	[0.027]	[0.036]	[0.026]	[0.076]	[0.033]
Conscientiousnes	0.43	-11.587*	0.098†	0.103†	0.002	0.015	-0.086	0.002	1.17	-2.344	0.007	0.003	-0.006	-0.012	0.028
	0.787	[5.080]	[0.051]	[0.053]	[0.011]	[0.090]	[0.058]	[0.007]	0.468	[2.909]	[0.028]	[0.033]	[0.022]	[0.084]	[0.041]
Emotional Stability	1.39	-12.435*	0.072	0.07	-0.011	0.146†	0.017	-0.004	0.84	-7.356†	0.021	0.008	-0.016	0.108	-0.038
	0.249	[5.984]	[0.059]	[0.058]	[0.010]	[0.077]	[0.051]	[0.008]	0.549	[2.899]	[0.028]	[0.034]	[0.019]	[0.073]	[0.05]
Intellect/Imagination	0.43	-5.204	0.039	0.046	0.01	0.116	0.031	0.001	0.90	-1.056	-0.004	-0.024	-0.022	0.051	-0.028
	0.788	[6.564]	[0.066]	[0.064]	[0.011]	[0.083]	[0.061]	[0.008]	0.522	[2.993]	[0.029]	[0.041]	[0.026]	[0.082]	[0.045]
Locus of control	0.81	1.266	0.000	-0.008	0.011	-0.064	-0.003	-0.017*	0.99	3.446	-0.023	-0.037	-0.012	-0.107	-0.018
	0.524	[4.963]	[0.052]	[0.052]	[0.010]	[0.077]	[0.051]	[0.006]	0.506	[3.505]	[0.035]	[0.042]	[0.016]	[0.085]	[0.044]
Intensity seeking	1.10	3.390*	-0.057	-0.076	-0.007	0.026	-0.003	0.000	0.96	1.534	0.015	0.000	-0.009	0.025	0.023
	0.366	[1.423]	[0.061]	[0.061]	[0.012]	[0.027]	[0.020]	[0.010]	0.509	[0.787]	[0.029]	[0.041]	[0.024]	[0.032]	[0.016]
Novelty seeking	0.46	2.03	-0.075	-0.085	-0.044*	-0.029	-0.003	-0.006	1.69	0.432	0.009	-0.028	-0.032	-0.02	0.016
	0.762	[1.423]	[0.068]	[0.068]	[0.018]	[0.027]	[0.025]	[0.014]	0.256	[0.816]	[0.032]	[0.04]	[0.023]	[0.023]	[0.017]

* Significant at 5%, †Significant at 10%. Robust standard errors in brackets. Clustered at school level.
** Means of 130 simulations which randomly selected five friends from a sample of people (real friends and respondent not included) who attended the same school and academic year as the respondent. Four of them are of the same gender and one of the opposite gender. Friend E is always the opposite gender friend.

Table 11: The dynamics of friendship and homophily. OLS results.

Characteristic	Age at measure (years)	Mean of var Resp	Friends	N	β (1)	Female (2)	GG (3)	GB (4)	BG (5)	Same school (8)	Mean of var Resp	Friends	N	β (9)	Female (10)	GG (11)	GB (12)	BG (13)	Same school (16)
KS1 score	7	105.7	105.1	1,170	0.273* [0.053]	-6.079 [6.584]	0.073 [0.063]	0.083 [0.063]	0.001 [0.010]	-0.002 [0.005]	106.0	104.4	3,563	0.248* [0.033]	3.412 [3.980]	-0.024 [0.038]	-0.017 [0.038]	0.013* [0.005]	0 [0.004]
KS2 score	11	104.6	104.4	2,498	0.386* [0.043]	2.118 [5.284]	-0.012 [0.050]	-0.01 [0.050]	0 [0.007]	0.004 [0.003]	105.5	104.2	2,449	0.326* [0.045]	-0.667 [5.446]	0.004 [0.051]	0.007 [0.051]	0.007 [0.005]	-0.003 [0.005]
KS3 score	14	106.1	105.2	3,929	0.383* [0.033]	-5.489 [3.884]	0.052 [0.037]	0.056 [0.037]	0.009* [0.005]	0.011* [0.004]	106.2	103.8	440	0.215† [0.125]	-19.337 [14.066]	0.176 [0.132]	0.18 [0.133]	0.013 [0.012]	-0.003 [0.007]
Total IQ	8	107.4	107.5	791	0.248* [0.064]	-0.838 [8.278]	0.021 [0.076]	0.055 [0.077]	0.003 [0.028]	-0.032* [0.011]	108.4	107.8	1,949	0.191* [0.043]	-8.567 [5.346]	0.064 [0.049]	0.052 [0.049]	0.011 [0.012]	0.006 [0.010]
Total difficulties score	7	97.4	98.3	569	0.028 [0.054]	-6.839 [7.008]	0.064 [0.070]	0.057 [0.070]	0.028 [0.020]	0.006 [0.008]	98.2	98.4	1,800	0.094* [0.047]	-0.093 [5.235]	-0.017 [0.053]	-0.005 [0.053]	-0.001 [0.009]	0.005 [0.008]
Total difficulties score	12	97.8	97.8	1,867	0.095* [0.039]	-2.467 [4.776]	0.013 [0.049]	0.01 [0.049]	-0.016† [0.009]	-0.003 [0.006]	98.2	98.5	961	0.142* [0.065]	3.109 [7.397]	-0.04 [0.076]	-0.053 [0.076]	-0.013 [0.010]	-0.019* [0.009]
Total difficulties score	13	97.9	97.7	1,712	0.083 [0.055]	3.925 [6.028]	-0.057 [0.062]	-0.059 [0.062]	-0.003 [0.011]	-0.004 [0.007]	98.4	98.9	462	0.025 [0.061]	5.279 [8.111]	-0.043 [0.083]	-0.057 [0.084]	-0.002 [0.013]	-0.017 [0.012]
Locus of control	8	99.6	99.5	611	0.096 [0.060]	-5.864 [7.544]	0.056 [0.076]	0.05 [0.077]	-0.006 [0.025]	0.017† [0.009]	98.9	98.9	1,474	0.133* [0.049]	4.308 [5.619]	-0.043 [0.057]	-0.023 [0.058]	-0.004 [0.010]	0.001 [0.010]
Total self esteem	8	100.8	101.3	719	0.174* [0.082]	14.631 [9.511]	-0.136 [0.094]	-0.138 [0.094]	-0.028 [0.022]	-0.024* [0.008]	100.8	100.6	1,794	0.104† [0.055]	15.006* [6.224]	-0.152* [0.061]	-0.156* [0.061]	0.002 [0.010]	-0.006 [0.008]
Child makes friends easily	9	0.7	0.8	932	0.049 [0.067]	0.057 [0.064]	-0.057 [0.074]	-0.025 [0.096]	-0.102 [0.104]	0.052 [0.041]	0.7	0.7	2,050	0.019 [0.059]	-0.011 [0.044]	0.061 [0.054]	0.027 [0.062]	0.014 [0.047]	-0.056 [0.039]
Child has lots of friends	9	0.9	0.9	926	0.082 [0.066]	0.096 [0.071]	-0.109 [0.075]	-0.015 [0.077]	-0.073 [0.071]	-0.024 [0.023]	0.9	0.9	2,074	0.011 [0.056]	0.028 [0.050]	-0.024 [0.054]	-0.095 [0.059]	-0.003 [0.032]	0.015 [0.030]
Child has more friends than most other kids	9	0.4	0.4	880	0.095 [0.073]	-0.013 [0.047]	-0.087 [0.072]	-0.144 [0.121]	-0.004 [0.145]	-0.006 [0.058]	0.4	0.4	1,942	0.222* [0.069]	0.047 [0.032]	-0.192* [0.056]	-0.196* [0.068]	-0.127† [0.070]	-0.085 [0.055]

Continued on next page

Table 11: The dynamics of friendship and homophily. OLS results. Continued

	Age at measure (years)	Friends at the time of the measure								Future Friends								
		Mean of var		N	β	Female	Gender composition			Same school	Mean of var		N	β	Female	Gender composition			Same school
Characteristic		Resp	Friends		(1)	(2)	GG (3)	GB (4)	BG (5)	(8)	Resp	Friends		(9)	(10)	GG (11)	GB (12)	BG (13)	(16)
Fat mass	9	100.7	99.5	878	0.142 [0.098]	5.993 [10.375]	-0.016 [0.108]	-0.028 [0.109]	-0.008 [0.020]	-0.012 [0.008]	100.6	101.0	1,935	-0.011 [0.039]	-1.915 [4.378]	0.074† [0.045]	0.080† [0.045]	0.011 [0.009]	-0.005 [0.008]
Fat percentage	11	99.2	99.0	1,208	0.011 [0.052]	-5.374 [6.325]	0.045 [0.063]	0.022 [0.064]	-0.005 [0.014]	-0.007 [0.006]	99.2	100.2	1,163	0.008 [0.056]	-3.014 [6.380]	0.019 [0.064]	0.033 [0.064]	-0.004 [0.012]	0.008 [0.010]
Fat percentage	12	99.3	99.4	1,787	0.035 [0.052]	-5.457 [5.720]	0.048 [0.057]	0.047 [0.057]	-0.005 [0.011]	-0.005 [0.006]	99.9	100.4	907	-0.015 [0.071]	-9.159 [7.736]	0.087 [0.078]	0.091 [0.078]	0.006 [0.013]	0.01 [0.010]
Fat percentage	13	99.2	99.7	2,017	0.078† [0.044]	-5.276 [4.925]	0.048 [0.049]	0.041 [0.050]	0.001 [0.009]	-0.012† [0.006]	99.5	99.6	531	0.059 [0.083]	-10.209 [9.817]	0.105 [0.100]	0.097 [0.101]	0.007 [0.015]	0.015 [0.012]
Fat percentage	15	99.5	99.5	2,172	0.049 [0.057]	-7.72 [6.255]	0.083 [0.064]	0.08 [0.064]	0.009 [0.009]	-0.014* [0.007]	99.2	99.8	71	-0.253 [0.185]	-45.377† [24.675]	0.413† [0.242]	0.423† [0.239]	-0.096† [0.055]	0.037† [0.020]
Plays a musical instrument	13	0.6	0.6	2,264	0.173* [0.051]	0.032 [0.035]	0.059 [0.049]	0.116† [0.068]	0.118† [0.061]	-0.089* [0.034]	0.6	0.6	618	0.023 [0.109]	0.037 [0.066]	0.111 [0.104]	0.106 [0.113]	0.093 [0.097]	0.013 [0.073]
Arrives late to lessons (often/sometimes=1, never=0)	15	0.3	0.3	2,349	0.118† [0.066]	0.075* [0.023]	0.113* [0.055]	0.066 [0.077]	0.124 [0.081]	-0.067 [0.052]	0.4	0.3	75	0.393† [0.229]	0.032 [0.118]	-0.033 [0.234]	-0.234 [0.271]	-0.337† [0.179]	0.032 [0.203]
Skips classes (often/sometimes=1, never=0)	15	0.1	0.1	2,363	0.04 [0.086]	-0.006 [0.015]	0.155* [0.059]	0.170† [0.089]	0.15 [0.099]	-0.05 [0.078]	0.2	0.2	77	0.723* [0.319]	0.041 [0.096]	-0.789* [0.201]	-0.477† [0.250]	-0.669† [0.339]	-0.132 [0.292]
likelihood goes to uni (very= 2, unsure=1, none=0)	15	1.7	1.7	2,418	0.211* [0.045]	0.029 [0.091]	0.004 [0.051]	0.012 [0.053]	0.017 [0.025]	0.004 [0.017]	1.5	1.5	76	0.233 [0.181]	-0.682† [0.397]	0.18 [0.225]	0.195 [0.232]	-0.351* [0.101]	0 [0.096]
Intensity seeking	11	24.1	24.5	1,325	0.111* [0.049]	-2.414 [1.552]	0.004 [0.060]	-0.004 [0.060]	0.023 [0.026]	0.006 [0.009]	24.2	24.5	1,373	0.150* [0.042]	-2.603* [1.324]	0.006 [0.052]	-0.015 [0.051]	-0.004 [0.017]	-0.065* [0.015]
Intensity seeking	14	25.3	25.7	2,233	0.137* [0.038]	-1.505 [1.215]	-0.017 [0.045]	-0.006 [0.044]	0.042* [0.016]	-0.004 [0.010]	25.8	26.5	239	0.191† [0.099]	1.868 [3.221]	-0.116 [0.125]	-0.144 [0.119]	0.025 [0.032]	-0.056* [0.023]

* Significant at 5%. †Significant at 10%. Robust standard errors in brackets.

Table 12: The dynamics of friendship and homophily. School Fixed Effects.

				Friends at the time of the measure									Future Friends						
	Age at	Mean of var		N	β	Female	Gender composition			Same school	Mean of var		N	β	Female	Gender composition			Same school
Characteristic	measure						GG	GB	BG							GG	GB	BG	
	(years)	Resp	Friends	N_{Sch}	(1)	(2)	(3)	(4)	(5)	(8)	Resp	Friends	N_{Sch}	(9)	(10)	(11)	(12)	(13)	(16)
KS1 score	7	105.7	105.2	1,143	0.212*	-7.619	0.086	0.099	0.003	0.002	106.0	104.5	3,468	0.192*	2.026	-0.012	-0.004	0.009*	0.001
				72	[0.052]	[8.074]	[0.079]	[0.077]	[0.010]	[0.006]			61	[0.045]	[5.033]	[0.048]	[0.049]	[0.004]	[0.005]
KS2 score	11	104.6	104.4	2,395	0.309*	-1.507	0.021	0.025	0.002	0.006†	105.4	104.2	2,374	0.255*	-2.599	0.022	0.025	0.005	0
				76	[0.045]	[5.669]	[0.055]	[0.055]	[0.007]	[0.003]			62	[0.055]	[5.412]	[0.051]	[0.052]	[0.005]	[0.004]
KS3 score	14	106.1	105.2	3,878	0.292*	-8.445*	0.078†	0.084*	0.008†	0.013*	106.1	103.7	436	0.063	-27.202	0.251	0.262	0.014	-0.001
				55	[0.035]	[4.157]	[0.040]	[0.040]	[0.004]	[0.005]			50	[0.167]	[17.285]	[0.163]	[0.162]	[0.012]	[0.008]
Total IQ	8	107.5	107.4	747	0.201*	-2.988	0.046	0.071	0.017	-0.026*	108.3	107.6	1,823	0.126*	-9.199†	0.065	0.052	0.008	0.013
				67	[0.065]	[7.897]	[0.072]	[0.071]	[0.036]	[0.011]			58	[0.039]	[4.613]	[0.042]	[0.044]	[0.010]	[0.012]
Total difficulties score	7	97.4	98.2	529	-0.07	-12.922	0.124	0.117	0.047	-0.001	98.3	98.4	1,684	0.009	-5.535	0.039	0.044	-0.003	0.005
				64	[0.069]	[7.777]	[0.079]	[0.076]	[0.030]	[0.009]			61	[0.052]	[4.743]	[0.049]	[0.049]	[0.009]	[0.008]
Total difficulties score	12	97.8	97.8	1,715	0.055	-5.634	0.044	0.037	-0.014	-0.005	98.2	98.7	903	0.079	-0.639	0.005	-0.009	-0.012	-0.017†
				61	[0.038]	[6.282]	[0.061]	[0.061]	[0.014]	[0.007]			58	[0.065]	[8.459]	[0.083]	[0.085]	[0.012]	[0.010]
Total difficulties score	13	97.9	97.7	1,575	0.025	-0.996	-0.008	-0.006	-0.002	-0.005	98.5	99.0	429	0.014	17.631	-0.171†	-0.194†	-0.003	-0.014
				59	[0.050]	[6.029]	[0.060]	[0.060]	[0.012]	[0.007]			51	[0.064]	[10.526]	[0.100]	[0.104]	[0.017]	[0.015]
Locus of control	8	99.7	99.6	578	0.087	-5.866	0.054	0.045	-0.027	0.009	98.9	99.0	1,372	0.091	1.742	-0.023	-0.007	-0.002	-0.004
				64	[0.073]	[9.117]	[0.091]	[0.095]	[0.034]	[0.012]			57	[0.059]	[6.714]	[0.066]	[0.065]	[0.011]	[0.013]
Total self esteem	8	100.8	101.3	673	0.167	16.581	-0.152	-0.164	-0.039†	-0.019*	100.7	100.7	1,681	0.091	15.037†	-0.154*	-0.155*	0.004	-0.005
				66	[0.133]	[14.244]	[0.143]	[0.144]	[0.023]	[0.009]			58	[0.072]	[7.736]	[0.075]	[0.076]	[0.009]	[0.011]
Child makes friends easily	9	0.7	0.8	883	0.046	0.07	-0.089	-0.05	-0.114	0.035	0.7	0.7	1,918	0.008	0.003	0.058	0.027	0.037	-0.05
				70	[0.076]	[0.072]	[0.091]	[0.095]	[0.109]	[0.044]			60	[0.066]	[0.059]	[0.066]	[0.079]	[0.052]	[0.032]
Child has lots of friends	9	0.9	0.9	877	0.067	0.119	-0.13	-0.068	-0.096	-0.034	0.9	0.9	1,941	-0.019	0.013	-0.009	-0.072	-0.013	0.012
				70	[0.084]	[0.077]	[0.085]	[0.087]	[0.080]	[0.021]			60	[0.067]	[0.067]	[0.064]	[0.071]	[0.032]	[0.043]
Child has more friends than most other kids	9	0.4	0.4	831	0.127†	-0.003	-0.134†	-0.122	-0.013	-0.019	0.4	0.4	1,816	0.191*	0.037	-0.169*	-0.157*	-0.077	-0.107*
				70	[0.073]	[0.059]	[0.080]	[0.142]	[0.137]	[0.060]			60	[0.076]	[0.043]	[0.062]	[0.073]	[0.067]	[0.047]

Table 12: The dynamics of friendship and homophily. School Fixed Effects. Continued

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N N_Sch	β (1)	Female (2)	GG (3)	GB (4)	BG (5)	Same school (8)	Mean of var Resp	Mean of var Friends	N N_Sch	β (9)	Female (10)	GG (11)	GB (12)	BG (13)	Same school (16)
							Friends at the time of the measure (Gender composition)									**Future Friends** (Gender composition)			
Fat mass	9	100.6	99.5	828	0.08	3.398	0.014	-0.002	-0.001	-0.011	100.6	101.0	1,831	-0.051	-4.752	0.101†	0.104†	0.015	-0.011
				65	[0.099]	[9.749]	[0.100]	[0.102]	[0.016]	[0.008]			58	[0.049]	[5.329]	[0.053]	[0.056]	[0.011]	[0.010]
Fat percentage	11	99.1	98.9	1,120	-0.067	-9.305	0.094	0.063	-0.005	-0.006	99.3	100.2	1,103	-0.048	-3.341	0.026	0.037	-0.005	0.002
				68	[0.047]	[7.425]	[0.074]	[0.073]	[0.015]	[0.007]			56	[0.063]	[7.014]	[0.069]	[0.069]	[0.012]	[0.013]
Fat percentage	12	99.3	99.5	1,657	0.003	-6.626	0.065	0.058	-0.005	-0.01	100.0	100.4	866	-0.066	-9.605	0.087	0.093	0.006	0.007
				61	[0.066]	[7.747]	[0.077]	[0.078]	[0.012]	[0.006]			59	[0.069]	[8.003]	[0.081]	[0.080]	[0.011]	[0.010]
Fat percentage	13	99.2	99.8	1,874	0.021	-7.115	0.071	0.061	0.008	-0.019*	99.7	99.7	502	0.011	-11.348	0.108	0.103	0.006	0.003
				61	[0.048]	[5.908]	[0.058]	[0.058]	[0.008]	[0.007]			53	[0.112]	[10.355]	[0.110]	[0.110]	[0.018]	[0.014]
Fat percentage	15	99.5	99.6	2,021	0.009	-6.722	0.076	0.068	0.011	-0.019*	99.5	100.0	68	-0.161	-36.75	0.333	0.369	-0.041	0.02
				59	[0.063]	[6.877]	[0.066]	[0.068]	[0.009]	[0.007]			35	[0.531]	[57.942]	[0.586]	[0.594]	[0.053]	[0.035]
Plays a musical instrument	13	0.6	0.6	2,090	0.078	0.013	0.044	0.127†	0.129*	-0.064	0.6	0.6	581	-0.114	0.052	0.151	0.173	0.175†	0.02
				62	[0.054]	[0.048]	[0.062]	[0.069]	[0.063]	[0.040]			55	[0.106]	[0.077]	[0.098]	[0.122]	[0.097]	[0.073]
Arrives late to lessons (often/sometimes=1, never=0)	15	0.3	0.3	2,175	0.125*	0.117*	0.077	0.047	0.1	-0.093	0.4	0.3	71	0.251	-0.153	0.034	0	-1.078*	0.267
				60	[0.059]	[0.034]	[0.056]	[0.067]	[0.064]	[0.068]			34	[0.510]	[0.150]	[0.530]	[0.000]	[0.458]	[0.372]
Skips classes (often/sometimes=1, never=0)	15	0.1	0.1	2,188	0.042	0.021	0.165†	0.152	0.128	-0.077	0.2	0.2	73	0.104	0.021	0	-0.104	-0.990*	0.25
				60	[0.093]	[0.020]	[0.089]	[0.120]	[0.111]	[0.079]			34	[0.130]	[0.062]	[0.000]	[0.149]	[0.171]	[0.220]
likelihood goes to uni (very= 2, unsure=1, none=0)	15	1.7	1.7	2,236	0.121*	0.022	-0.002	0.019	0.014	0.017	1.5	1.5	72	-0.126	-1.238*	0.302	0.245	-0.439†	0.084
				60	[0.051]	[0.116]	[0.064]	[0.065]	[0.025]	[0.018]			33	[0.198]	[0.410]	[0.227]	[0.205]	[0.258]	[0.168]
Intensity seeking	11	24.1	24.5	1,241	0.034	-2.653	0.012	0.018	0.013	0.005	24.3	24.5	1,295	0.106*	-3.269*	0.03	0.007	-0.029	-0.075*
				69	[0.056]	[1.925]	[0.074]	[0.075]	[0.024]	[0.012]			56	[0.039]	[1.324]	[0.055]	[0.049]	[0.017]	[0.018]
Intensity seeking	14	25.3	25.7	2,082	0.088*	-1.382	-0.025	-0.013	0.030†	-0.002	25.8	26.4	224	0.146	1.645	-0.087	-0.107	0.005	-0.088*
				61	[0.041]	[1.311]	[0.048]	[0.049]	[0.016]	[0.013]			50	[0.102]	[2.781]	[0.112]	[0.108]	[0.036]	[0.037]

* Significant at 5%. †Significant at 10%

Robust standard errors in brackets. Clustered at school level.

Table 13: Factors affecting the degree of homophily. OLS results.

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N	β (1)	Female (2)	Gender composition GG (3)	GB (4)	BG (5)	Same school (6)	Length of friendship (years) 0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	Closeness quintile[1/] first (11)	second (12)	third (13)	fourth (14)
KS1 score	7	106.0	105.3	1,043	0.258* [0.050]	-9.698 [6.493]	0.103† [0.062]	0.110† [0.062]	0.002 [0.010]	0.001 [0.005]				-0.004 [0.004]	-0.017* [0.007]	0.003 [0.007]	0.003 [0.006]	0.003 [0.006]
KS2 score	11	104.8	104.5	2,240	0.365* [0.043]	0.787 [5.253]	0.000 [0.050]	0.000 [0.050]	0.004 [0.008]	0.002 [0.004]		0.015 [0.011]	-0.011* [0.004]	-0.004 [0.004]	-0.026* [0.005]	-0.008* [0.004]	-0.004 [0.004]	-0.001 [0.004]
KS3 score	14	106.3	105.3	3,585	0.354* [0.033]	-7.683* [3.905]	0.072* [0.037]	0.075* [0.037]	0.009† [0.005]	0.013* [0.004]	0.009† [0.005]	0.003 [0.004]	-0.010* [0.004]	-0.006 [0.004]	-0.015* [0.004]	-0.004 [0.003]	-0.002 [0.003]	0.000 [0.003]
Total IQ	8	108.1	107.6	715	0.234* [0.058]	-3.806 [7.757]	0.036 [0.071]	0.065 [0.073]	0.000 [0.030]	-0.026* [0.012]				-0.014 [0.011]	-0.042* [0.017]	-0.001 [0.015]	-0.01 [0.015]	0.015 [0.015]
Total IQ	16	51.1	51.0	1,981	0.269* [0.045]	-2.989 [2.533]	0.032 [0.048]	0.024 [0.048]	0.008 [0.018]	0.014 [0.013]	0.003 [0.016]	0.02 [0.015]	-0.060* [0.016]	-0.021 [0.015]	-0.024 [0.015]	0.007 [0.013]	0.015 [0.013]	0.027* [0.013]
Total difficulties score	7	97.4	98.2	520	0.007 [0.062]	-5.93 [7.775]	0.056 [0.078]	0.049 [0.078]	0.027 [0.021]	0.009 [0.008]				0.01 [0.007]	0.014 [0.013]	0.003 [0.011]	0.009 [0.012]	0.003 [0.011]
Total difficulties score	12	97.8	97.8	1,716	0.069 [0.044]	-3.122 [5.187]	0.016 [0.053]	0.015 [0.053]	-0.014 [0.010]	-0.004 [0.006]		0.001 [0.007]	0.005 [0.007]	0.004 [0.006]	0.030* [0.008]	0.009 [0.007]	0.009 [0.006]	0.003 [0.006]
Total difficulties score	13	97.9	97.7	1,566	0.081 [0.059]	4.949 [6.389]	-0.071 [0.065]	-0.073 [0.065]	0.000 [0.011]	-0.008 [0.007]	0.031 [0.035]	0.015* [0.007]	0.012† [0.007]	0.011† [0.007]	0.023* [0.008]	0.001 [0.006]	0.001 [0.006]	0.002 [0.006]
Locus of control	8	99.3	99.7	552	0.076 [0.065]	-6.359 [7.873]	0.067 [0.079]	0.063 [0.080]	-0.002 [0.027]	0.016 [0.010]				0.009 [0.009]	0.028* [0.014]	0.003 [0.014]	0.008 [0.013]	-0.006 [0.015]
Locus of control	17	98.9	99.1	1,954	0.089† [0.050]	0.771 [5.503]	0.001 [0.056]	-0.007 [0.057]	0.009 [0.010]	-0.009 [0.007]	-0.001 [0.009]	-0.006 [0.009]	0.002 [0.009]	-0.007 [0.009]	0.020* [0.007]	0.005 [0.007]	0.017* [0.007]	0.003 [0.007]
Total self esteem	8	100.7	101.5	650	0.151† [0.090]	12.727 [10.207]	-0.113 [0.100]	-0.121 [0.101]	-0.040† [0.023]	-0.025* [0.009]				0.001 [0.008]	-0.02 [0.012]	0.011 [0.012]	-0.017 [0.012]	0.005 [0.011]

Table 13: Factors affecting the degree of homophily. OLS results. Continued

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N	β (1)	Female (2)	Gender composition GG (3)	GB (4)	BG (5)	Same school (6)	Length of friendship (years) 0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	Closeness quintile[1/] first (11)	second (12)	third (13)	fourth (14)
Child makes friends easily	9	0.7	0.8	839	0.109 [0.081]	0.089 [0.068]	-0.097 [0.079]	-0.074 [0.101]	-0.153 [0.116]	0.048 [0.044]				-0.017 [0.037]	-0.077 [0.056]	-0.047 [0.055]	0.032 [0.049]	-0.034 [0.053]
Child has lots of friends	9	0.9	0.9	834	0.135† [0.075]	0.114 [0.078]	-0.129 [0.082]	-0.039 [0.083]	-0.1 [0.082]	-0.024 [0.025]				-0.028 [0.023]	-0.049 [0.038]	-0.001 [0.034]	-0.006 [0.032]	0.001 [0.033]
Child has more friends than most other kids	9	0.4	0.4	790	0.191† [0.097]	-0.02 [0.051]	-0.078 [0.078]	-0.117 [0.125]	-0.031 [0.141]	-0.011 [0.062]				-0.04 [0.057]	-0.171* [0.081]	-0.09 [0.088]	-0.108 [0.081]	-0.098 [0.087]
Fat mass	9	100.8	99.6	786	0.232* [0.108]	14.538 [11.258]	-0.102 [0.117]	-0.115 [0.119]	-0.022 [0.020]	-0.014 [0.009]				0.01 [0.008]	-0.002 [0.011]	-0.005 [0.012]	-0.006 [0.012]	-0.014 [0.011]
Fat percentage	11	99.1	99.0	1,089	0.033 [0.058]	-4.619 [6.695]	0.037 [0.067]	0.012 [0.068]	-0.007 [0.014]	-0.005 [0.008]		-0.008 [0.027]	-0.001 [0.009]	0.002 [0.008]	-0.003 [0.010]	-0.006 [0.010]	-0.007 [0.010]	-0.006 [0.009]
Fat percentage	12	99.2	99.5	1,631	0.068 [0.057]	-1.526 [6.068]	0.012 [0.061]	0.01 [0.061]	-0.002 [0.012]	-0.007 [0.007]		0.001 [0.008]	-0.002 [0.007]	0.005 [0.007]	0.01 [0.008]	0.006 [0.007]	-0.005 [0.007]	-0.001 [0.007]
Fat percentage	13	99.2	99.8	1,854	0.103* [0.047]	-3.246 [5.221]	0.029 [0.052]	0.019 [0.053]	0.000 [0.009]	-0.016* [0.007]	0.02 [0.017]	0.002 [0.007]	0.007 [0.008]	0.005 [0.007]	0.000 [0.007]	-0.003 [0.007]	-0.002 [0.007]	-0.008 [0.007]
Fat percentage	15	99.4	99.5	2,007	0.088 [0.065]	-5.39 [6.744]	0.061 [0.069]	0.054 [0.069]	0.006 [0.009]	-0.014† [0.008]	0.000 [0.009]	-0.006 [0.008]	-0.001 [0.009]	0.000 [0.009]	0.001 [0.007]	-0.006 [0.007]	-0.011 [0.007]	-0.009 [0.007]
Plays a musical instrument	13	0.6	0.6	2,089	0.212* [0.063]	0.02 [0.038]	0.087† [0.052]	0.149* [0.071]	0.161* [0.063]	-0.051 [0.039]	-0.011 [0.149]	-0.071† [0.042]	-0.119* [0.044]	-0.002 [0.039]	-0.094* [0.043]	-0.054 [0.042]	-0.082* [0.040]	0.001 [0.040]
Extraversion	14	99.8	100.7	2,010	0.234* [0.043]	4.952 [5.085]	-0.022 [0.051]	-0.013 [0.051]	0.023* [0.009]	0.003 [0.007]	-0.029* [0.009]	-0.021* [0.007]	-0.024* [0.008]	-0.021* [0.007]	-0.058* [0.007]	-0.033* [0.007]	-0.021* [0.006]	-0.009 [0.006]
Agreeableness	14	102.0	102.0	1,934	0.003 [0.047]	5.15 [5.286]	0.001 [0.053]	0.01 [0.054]	0.016 [0.010]	-0.002 [0.006]	-0.015† [0.008]	-0.017* [0.007]	-0.008 [0.007]	-0.01 [0.006]	-0.021* [0.006]	-0.012† [0.006]	-0.019* [0.006]	-0.011† [0.006]
Conscientiousnes	14	101.9	101.1	1,852	0.022 [0.043]	-12.477* [5.043]	0.110* [0.050]	0.115* [0.050]	0.006 [0.009]	0.002 [0.008]	-0.004 [0.010]	-0.004 [0.008]	0.007 [0.009]	0.001 [0.008]	0.002 [0.007]	0.002 [0.007]	0.000 [0.007]	0.000 [0.007]
Emotional Stability	14	100.4	100.3	1,882	-0.006 [0.041]	-11.824* [4.995]	0.069 [0.049]	0.067 [0.049]	-0.009 [0.009]	0.005 [0.007]	-0.020* [0.010]	-0.012 [0.008]	-0.018* [0.008]	-0.009 [0.007]	-0.007 [0.007]	-0.005 [0.007]	0.001 [0.007]	-0.002 [0.007]
Intellect/Imagination	14	101.0	101.0	1,934	0.123* [0.040]	-8.564† [4.954]	0.073 [0.049]	0.078 [0.049]	0.016† [0.009]	0.006 [0.007]	-0.009 [0.010]	-0.009 [0.008]	-0.025* [0.009]	-0.020* [0.008]	-0.026* [0.007]	-0.021* [0.007]	-0.024* [0.007]	-0.011 [0.007]

Table 13: Factors affecting the degree of homophily. OLS results. Continued

Characteristic	Age at measure (years)	Mean of var		N	β (1)	Female (2)	Gender composition			Same school (6)	Length of friendship (years)				Closeness quintile[1/]			
		Resp	Friends				GG (3)	GB (4)	BG (5)		0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	first (11)	second (12)	third (13)	fourth (14)
Arrives late to lessons (often/sometimes=1, never=0)	15	0.3	0.3	2,175	0.216* [0.087]	0.080* [0.024]	0.112† [0.058]	0.04 [0.079]	0.122 [0.085]	-0.023 [0.058]	0.015 [0.071]	-0.047 [0.068]	-0.115 [0.071]	-0.059 [0.066]	-0.192* [0.060]	-0.165* [0.058]	-0.086 [0.057]	-0.065 [0.058]
Skips classes (often/sometimes=1, never=0)	15	0.1	0.1	2,194	-0.044 [0.111]	-0.006 [0.016]	0.162* [0.065]	0.153 [0.096]	0.051 [0.096]	-0.081 [0.081]	0.180† [0.094]	0.091 [0.090]	0.101 [0.093]	0.055 [0.086]	0.036 [0.081]	0.065 [0.079]	0.009 [0.073]	-0.031 [0.078]
likelihood goes to uni (very= 2, unsure=1, none=0)	15	1.7	1.7	2,237	0.231* [0.050]	0.005 [0.094]	0.021 [0.053]	0.035 [0.055]	0.012 [0.028]	0.006 [0.019]	-0.024 [0.026]	-0.006 [0.023]	-0.009 [0.024]	0.005 [0.022]	-0.073* [0.020]	-0.040* [0.018]	-0.038* [0.017]	-0.027 [0.017]
Intensity seeking	11	24.0	24.4	1,201	0.137* [0.051]	-1.476 [1.551]	-0.026 [0.060]	-0.027 [0.061]	0.045 [0.028]	0.016 [0.013]		0.073 [0.087]	0.01 [0.015]	-0.001 [0.013]	-0.035* [0.016]	-0.055* [0.015]	-0.001 [0.015]	0.004 [0.015]
Intensity seeking	14	25.3	25.7	2,061	0.173* [0.041]	-1.163 [1.244]	-0.026 [0.046]	-0.015 [0.045]	0.040* [0.016]	-0.002 [0.011]	-0.013 [0.016]	-0.013 [0.013]	-0.028* [0.014]	-0.027* [0.013]	-0.070* [0.011]	-0.038* [0.011]	-0.027* [0.011]	-0.021† [0.011]
Intensity seeking	17	25.3	25.3	2,040	0.06 [0.051]	2.858* [1.397]	-0.038 [0.057]	-0.054 [0.059]	-0.013 [0.014]	0.005 [0.009]	0.011 [0.012]	0.008 [0.010]	0.000 [0.012]	0.016 [0.010]	0.020† [0.010]	0.018† [0.010]	0.002 [0.010]	0.006 [0.010]
Novelty seeking	17	23.0	22.9	2,005	0.110* [0.045]	2.370* [1.144]	-0.08 [0.049]	-0.103† [0.053]	-0.045* [0.015]	-0.017 [0.012]	0.039* [0.015]	0.014 [0.014]	0.039* [0.015]	0.02 [0.014]	0.046* [0.013]	0.038* [0.013]	0.016 [0.012]	0.02 [0.013]

* Significant at 5%. †Significant at 10%. Robust standard errors in brackets.
[1/] The least close friends are in the first quintile and the closest ones in the fifth quintile. Fifth quintile is the omitted category

Table 14: Factors affecting the degree of homophily. School fixed effects.

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N / N$_{Sch}$	β (1)	Female (2)	Gender composition GG (3)	Gender composition GB (4)	Gender composition BG (5)	Same school (6)	Length of friendship (years) 0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	Closeness quintile[1/] first (11)	second (12)	third (13)	fourth (14)
KS1 score	7	106.0	105.4	1,017 68	0.242* [0.052]	-6.909 [8.397]	0.078 [0.082]	0.085 [0.080]	0.002 [0.010]	0.006 [0.006]				-0.004 [0.005]	-0.014† [0.007]	0.007 [0.008]	0.005 [0.007]	0.007 [0.006]
KS2 score	11	104.8	104.5	2,140 73	0.327* [0.050]	1.444 [6.371]	-0.007 [0.062]	-0.005 [0.062]	0.002 [0.008]	0.006 [0.004]		0.024† [0.014]	-0.009† [0.004]	-0.003 [0.004]	-0.024* [0.005]	-0.006 [0.005]	-0.002 [0.005]	0.002 [0.004]
KS3 score	14	106.3	105.3	3,536 53	0.291* [0.038]	-7.956† [4.381]	0.074† [0.043]	0.078† [0.042]	0.007 [0.005]	0.015* [0.005]	0.011* [0.004]	0.005 [0.004]	-0.011* [0.004]	-0.003 [0.004]	-0.014* [0.004]	-0.003 [0.004]	-0.001 [0.004]	0.002 [0.003]
Total IQ	8	108.1	107.5	673 65	0.202* [0.067]	-3.739 [7.700]	0.039 [0.073]	0.06 [0.073]	0.011 [0.036]	-0.014 [0.011]				-0.016 [0.012]	-0.043* [0.018]	0.006 [0.018]	0.006 [0.020]	0.027 [0.019]
Total IQ	16	50.7	50.6	1,840 58	0.157* [0.054]	-2.924 [2.026]	0.034 [0.038]	0.032 [0.041]	0.019 [0.021]	0.035* [0.013]	-0.015 [0.017]	0.011 [0.015]	-0.072* [0.016]	-0.035* [0.016]	-0.022 [0.016]	0.01 [0.018]	0.021 [0.015]	0.033* [0.015]
Total difficulties score	7	97.4	98.1	483 62	-0.075 [0.086]	-9.502 [8.717]	0.094 [0.089]	0.082 [0.086]	0.053† [0.030]	0.002 [0.010]				0.015† [0.008]	0.01 [0.016]	-0.01 [0.012]	-0.001 [0.015]	-0.008 [0.011]
Total difficulties score	12	97.8	97.8	1,571 60	0.043 [0.042]	-4.654 [6.827]	0.029 [0.066]	0.026 [0.067]	-0.009 [0.015]	-0.009 [0.007]		0.000 [0.009]	0.008 [0.007]	0.006 [0.006]	0.031* [0.008]	0.009 [0.007]	0.009 [0.007]	0.006 [0.008]
Total difficulties score	13	98.0	97.7	1,435 58	0.032 [0.059]	0.447 [6.228]	-0.027 [0.062]	-0.025 [0.062]	0.001 [0.014]	-0.011 [0.009]	0.029 [0.034]	0.017* [0.007]	0.015† [0.009]	0.013† [0.008]	0.024* [0.011]	0.001 [0.007]	0.001 [0.006]	0.003 [0.008]
Locus of control	8	99.4	99.8	522 62	0.06 [0.084]	-9.173 [9.956]	0.095 [0.099]	0.09 [0.103]	-0.016 [0.036]	0.003 [0.014]				0.009 [0.011]	0.025 [0.018]	0.002 [0.021]	0.002 [0.016]	-0.012 [0.024]
Locus of control	17	99.0	99.2	1,797 59	0.06 [0.057]	1.927 [5.136]	-0.01 [0.053]	-0.02 [0.053]	0.011 [0.010]	-0.018* [0.006]	-0.005 [0.009]	-0.01 [0.009]	0.001 [0.008]	-0.012 [0.009]	0.022* [0.006]	0.008 [0.009]	0.020* [0.009]	0.003 [0.007]
Total self esteem	8	100.7	101.5	606 65	0.159 [0.148]	14.047 [16.089]	-0.123 [0.160]	-0.14 [0.163]	-0.060* [0.024]	-0.021* [0.009]				0.001 [0.008]	-0.018 [0.013]	0.016 [0.011]	-0.016 [0.012]	0.011 [0.010]

Continued on next page

Table 14: Factors affecting the degree of homophily. School fixed effects. Continued

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N / N_{Sch}	β (1)	Female (2)	GG (3)	GB (4)	BG (5)	Same school (6)	0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	first (11)	second (12)	third (13)	fourth (14)
Child makes friends easily	9	0.7	0.8	792	0.06	0.098	-0.112	-0.089	-0.175	0.016				0.013	-0.035	-0.02	0.062	-0.008
				66	[0.084]	[0.075]	[0.093]	[0.099]	[0.109]	[0.051]				[0.037]	[0.063]	[0.059]	[0.056]	[0.055]
Child has lots of friends	9	0.9	0.9	787	0.105	0.118	-0.128	-0.08	-0.131	-0.047†				-0.005	-0.056	-0.017	-0.02	-0.007
				66	[0.097]	[0.082]	[0.090]	[0.092]	[0.085]	[0.026]				[0.027]	[0.047]	[0.036]	[0.036]	[0.036]
Child has more friends than most other kids	9	0.4	0.4	743	0.211*	-0.005	-0.127†	-0.081	-0.063	-0.029				-0.011	-0.156†	-0.069	-0.128	-0.092
				66	[0.105]	[0.068]	[0.075]	[0.139]	[0.135]	[0.066]				[0.049]	[0.086]	[0.089]	[0.094]	[0.091]
Fat mass	9	100.7	99.5	738	0.171	11.407	-0.067	-0.084	-0.016	-0.015				0.008	-0.006	-0.008	-0.012	-0.019
				63	[0.110]	[10.562]	[0.110]	[0.113]	[0.019]	[0.011]				[0.010]	[0.012]	[0.014]	[0.017]	[0.014]
Fat percentage	11	99.1	98.9	1,006	-0.043	-8.046	0.081	0.044	-0.007	-0.006		-0.056	-0.006	0.001	-0.003	-0.005	-0.007	-0.003
				67	[0.056]	[7.440]	[0.074]	[0.074]	[0.016]	[0.011]		[0.046]	[0.010]	[0.011]	[0.015]	[0.010]	[0.012]	[0.009]
Fat percentage	12	99.2	99.5	1,508	0.024	-3.737	0.039	0.031	-0.001	-0.013		0.001	-0.002	0.006	0.011	0.006	-0.005	-0.001
				60	[0.073]	[8.027]	[0.079]	[0.080]	[0.013]	[0.008]		[0.009]	[0.008]	[0.007]	[0.007]	[0.008]	[0.007]	[0.007]
Fat percentage	13	99.2	99.8	1,719	0.046	-5.395	0.054	0.042	0.006	-0.023*	0.038*	0.000	0.006	0.003	0.000	-0.005	0.000	-0.011†
				60	[0.053]	[6.166]	[0.060]	[0.060]	[0.009]	[0.008]	[0.018]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]	[0.007]	[0.006]
Fat percentage	15	99.4	99.6	1,864	0.056	-4.022	0.049	0.036	0.007	-0.021*	0.004	-0.007	-0.001	0.002	-0.003	-0.011	-0.012	-0.014†
				59	[0.070]	[6.931]	[0.067]	[0.068]	[0.009]	[0.008]	[0.009]	[0.009]	[0.008]	[0.009]	[0.008]	[0.008]	[0.008]	[0.008]
Plays a musical instrument	13	0.6	0.6	1,924	0.132†	0.001	0.073	0.165*	0.175*	-0.024	-0.147	-0.088	-0.114†	-0.005	-0.116*	-0.081†	-0.079	-0.008
				61	[0.074]	[0.048]	[0.062]	[0.073]	[0.051]	[0.049]	[0.163]	[0.056]	[0.063]	[0.050]	[0.042]	[0.044]	[0.054]	[0.041]
Extraversion	14	99.8	100.6	1,872	0.204*	3.366	-0.004	0.002	0.022*	-0.01	-0.024*	-0.020*	-0.025*	-0.023*	-0.064*	-0.038*	-0.028*	-0.013†
				60	[0.047]	[7.335]	[0.073]	[0.072]	[0.010]	[0.006]	[0.009]	[0.007]	[0.008]	[0.008]	[0.010]	[0.010]	[0.008]	[0.006]
Agreeableness	14	102.1	101.9	1,793	-0.023	2.911	0.023	0.032	0.016	-0.006	-0.017*	-0.018*	-0.01	-0.012†	-0.019*	-0.011†	-0.016*	-0.008
				59	[0.056]	[6.099]	[0.061]	[0.062]	[0.011]	[0.007]	[0.007]	[0.006]	[0.009]	[0.007]	[0.008]	[0.006]	[0.006]	[0.007]
Conscientiousnes	14	101.9	101.1	1,716	-0.006	-9.292	0.076	0.082	0.001	-0.001	0.001	-0.001	0.014	0.007	0.005	0.002	0.004	0.004
				59	[0.045]	[5.779]	[0.058]	[0.059]	[0.011]	[0.007]	[0.009]	[0.009]	[0.010]	[0.010]	[0.008]	[0.007]	[0.008]	[0.007]
Emotional Stability	14	100.4	100.3	1,754	-0.04	-11.919†	0.068	0.063	-0.013	-0.001	-0.013	-0.013	-0.016	-0.009	-0.011	-0.008	-0.004	-0.007
				60	[0.051]	[6.490]	[0.065]	[0.063]	[0.011]	[0.009]	[0.010]	[0.008]	[0.010]	[0.008]	[0.007]	[0.008]	[0.008]	[0.008]
Intellect/Imagination	14	101.0	100.9	1,796	0.106*	-4.362	0.029	0.036	0.012	0.004	-0.017	-0.015	-0.031*	-0.027*	-0.028*	-0.024*	-0.024*	-0.007
				59	[0.049]	[7.141]	[0.071]	[0.070]	[0.011]	[0.009]	[0.013]	[0.011]	[0.009]	[0.012]	[0.009]	[0.008]	[0.009]	[0.008]

Table 14: Factors affecting the degree of homophily. School fixed effects. Continued

Characteristic	Age at measure (years)	Mean of var Resp	Mean of var Friends	N N_Sch	β (1)	Female (2)	Gender composition GG (3)	GB (4)	BG (5)	Same school (6)	Length of friendship (years) 0-2 (7)	3-4 (8)	5 (9)	6-12 (10)	Closeness quintile[1] first (11)	second (12)	third (13)	fourth (14)
Arrives late to lessons (often/sometimes=1, never=0)	15	0.3	0.3	2,011 60	0.225* [0.097]	0.135* [0.035]	0.076 [0.064]	0.021 [0.065]	0.109 [0.073]	-0.048 [0.067]	0.009 [0.064]	-0.07 [0.073]	-0.124* [0.060]	-0.069 [0.065]	-0.184* [0.084]	-0.160* [0.075]	-0.088 [0.060]	-0.057 [0.072]
Skips classes (often/sometimes=1, never=0)	15	0.1	0.1	2,029 60	-0.031 [0.116]	0.022 [0.021]	0.157† [0.086]	0.13 [0.125]	0.003 [0.130]	-0.101 [0.092]	0.191† [0.101]	0.062 [0.074]	0.105 [0.100]	0.038 [0.086]	0.023 [0.113]	0.074 [0.109]	0.011 [0.076]	-0.011 [0.106]
likelihood goes to uni (very= 2, unsure=1, none=0)	15	1.7	1.7	2,065 60	0.166* [0.051]	0.007 [0.110]	0.011 [0.060]	0.037 [0.063]	0.004 [0.026]	0.024 [0.022]	-0.033 [0.027]	-0.018 [0.023]	-0.019 [0.029]	-0.01 [0.026]	-0.075* [0.026]	-0.044* [0.021]	-0.042* [0.015]	-0.036* [0.018]
Intensity seeking	11	24.0	24.4	1,122 67	0.078 [0.058]	-1.265 [1.844]	-0.038 [0.069]	-0.023 [0.069]	0.037 [0.028]	0.018 [0.015]		0.05 [0.092]	0.016 [0.015]	0.001 [0.017]	-0.049* [0.022]	-0.059* [0.017]	-0.012 [0.017]	-0.002 [0.016]
Intensity seeking	14	25.3	25.7	1,916 60	0.128* [0.043]	-0.921 [1.328]	-0.037 [0.048]	-0.024 [0.049]	0.031† [0.016]	0.002 [0.015]	-0.02 [0.015]	-0.012 [0.014]	-0.032* [0.012]	-0.029* [0.012]	-0.065* [0.013]	-0.039* [0.012]	-0.027* [0.012]	-0.017* [0.009]
Intensity seeking	17	25.3	25.3	1,880 58	0.034 [0.051]	3.375* [1.456]	-0.058 [0.063]	-0.072 [0.061]	-0.008 [0.012]	-0.002 [0.012]	0.01 [0.015]	0.007 [0.013]	0.002 [0.014]	0.012 [0.012]	0.026† [0.013]	0.022† [0.013]	0.004 [0.013]	0.009 [0.012]
Novelty seeking	17	23.1	23.0	1,851 59	0.088 [0.053]	2.749† [1.394]	-0.105 [0.066]	-0.115† [0.064]	-0.045* [0.018]	-0.01 [0.015]	0.046* [0.015]	0.013 [0.018]	0.047* [0.018]	0.02 [0.015]	0.045* [0.019]	0.038* [0.017]	0.017 [0.013]	0.012 [0.014]

* Significant at 5%. †Significant at 10%

Robust standard errors in brackets. Clustered at school level.

[1] The least close friends are in the first quintile and the closest ones in the fifth quintile. Fifth quintile is the omitted category

Table 15: Homophily estimations. Individual fixed effects.

Characteristic	Mean of var Resp	Mean of var Friends	N N_i	β (1)	Gender composition GG (2)	Gender composition GB (3)	Gender composition BG (4)	Same school (5)	Length of friendship (years) 0-2 (6)	Length of friendship (years) 3-4 (7)	Length of friendship (years) 5 (8)	Length of friendship (years) 6-12 (9)	Closeness quintile[1] first (10)	Closeness quintile[1] second (11)	Closeness quintile[1] third (12)	fourth (13)
Basic Model																
KS scores	105.1	104.3	10,033	0.036*	0.011	0.011	0.000	0.001								
			2,117	[0.008]	[0.011]	[0.011]	[0.001]	[0.001]								
IQ	101.6	101.6	3,350	-0.004	0.009	0.011	-0.002	0.001								
			1,515	[0.019]	[0.024]	[0.024]	[0.003]	[0.003]								
Locus of control	99.1	99.1	3,038	0.034	-0.028	-0.031	0.001	0.002								
			1,494	[0.025]	[0.031]	[0.031]	[0.004]	[0.006]								
Intensity seeking	98.5	99.1	6,922	0.079*	0.003	-0.004	-0.011*	0.003								
			1,845	[0.027]	[0.031]	[0.031]	[0.006]	[0.003]								
Body fat percentage	99.4	99.6	9,351	0.011	-0.001	-0.002	-0.003	0.000								
			1,768	[0.010]	[0.012]	[0.012]	[0.002]	[0.002]								
Total difficulties score	97.8	98.0	5,293	-0.026†	0.02	0.02	0.002	0.004								
			1,613	[0.014]	[0.018]	[0.018]	[0.003]	[0.003]								
Adding length and closeness measures																
KS scores	105.4	104.5	9,044	0.033*	0.019†	0.019†	-0.001	0.001	0.002	0.003*	0.001	0.001	-0.001	-0.001	-0.002†	-0.001
			2,006	[0.008]	[0.011]	[0.011]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
IQ	101.9	101.7	3,069	0.001	0.011	0.012	0.000	0.001	-0.003	-0.002	-0.005	-0.003	0.003	0.000	0.000	0.000
			1,435	[0.019]	[0.024]	[0.024]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Locus of control	99.1	99.2	2,802	0.032	-0.012	-0.013	0.001	0.003	0.001	0.000	0.001	0.000	-0.002	-0.006	-0.003	-0.005
			1,411	[0.028]	[0.032]	[0.033]	[0.004]	[0.006]	[0.006]	[0.005]	[0.006]	[0.006]	[0.005]	[0.004]	[0.004]	[0.004]
Intensity seeking	98.4	99.1	6,353	0.082*	-0.006	-0.014	-0.008	0.002	-0.003	-0.002	-0.001	0.002	0.003	0.006	0.001	0.002
			1,740	[0.027]	[0.032]	[0.031]	[0.006]	[0.003]	[0.005]	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.003]	[0.003]
Body fat percentage	99.4	99.6	8,543	0.01	-0.002	-0.003	-0.004*	0.000	0.000	0.000	-0.001	-0.001	0.002	0.001	-0.001	0.003*
			1,669	[0.011]	[0.012]	[0.012]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]
Total difficulties score	97.8	98.0	4,841	-0.015	0.005	0.005	0.002	0.004	-0.023	-0.001	-0.003	-0.002	-0.002	0.002	0.002	0.000
			1,516	[0.015]	[0.019]	[0.019]	[0.003]	[0.003]	[0.016]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]

* Significant at 5%, †Significant at 10% . We also control for age of respondent and friend at the time of each measure and year (or round) fixed effects. Results are available upon request. [1]The least close friends are in the first quintile and the closest ones in the fifth quintile. Fifth quintile is the omitted category.

Appendix

A Friendship variables

Questions asked about Friendships
1. Are you happy with the number of friends you have got?
2. How often do you see your friends outside of school?
3. Do your friends understand you? (do they know what makes you happy or sad?)
4. Do you talk to your friends about problems?
5. Overall, how happy are you with your friends?

B Five Factor Model (FFM) - 50 statements.

Extraversion is defined as a *keen interest in other people and external events, and venturing with confidence into the unknown.* Agreeableness measures how compatible people are with other people, or how able they are to get along with others. Conscientiousness indicates how organised and aware of their environment a person is and how much they pay attention to details. Emotional stability is a dimension of personality defined *by stability and low anxiety at one end as opposed to instability and high anxiety at the other end.* Openness refers to how willing people are to make adjustments in notions and activities in accordance with new ideas or situations (The ALSPAC Study Team; 2010).

"You are going to see some statements describing people's behaviour. Please read each statement carefully and decide how well each statement describes you, from very like me to not at all like me. Describe yourself as you generally are now, not as you wish to be in the future."

Factor I (Surgency or Extraversion)
Am the life of the party
Don't talk a lot.
Feel comfortable around people.
Keep in the background.
Start conversations.
Have little to say.
Talk to a lot of different people at parties.
Don't like to draw attention to myself.
Don't mind being the centre of attention.
Am quiet around strangers.

Factor II (Agreeableness)
Feel little concern for others.
Am interested in people.
Insult people.
Sympathize with others' feelings.
Am not interested in other people's problems.
Have a soft heart.
Am not really interested in others.
Take time out for others.
Feel others' emotions.
Make people feel at ease.

Factor III (Conscientiousness)
Am always prepared.
Leave my belongings around.
Pay attention to details.
Make a mess of things.
Get household tasks done right away.
Often forget to put things back in their proper place.
Like order.
Avoid my duties.
Follow a plan.
Am exacting in my work.

Factor IV (Emotional Stability)
Get stressed out easily.
Am relaxed most of the time.
Worry about things.
Seldom feel sad.
Am easily disturbed.
Get upset easily.
Change my mood a lot.
Have frequent mood swings.
Get irritated easily.
Often feel sad.

Factor V (Intellect or Imagination)
Have a wide vocabulary.
Have difficulty understanding abstract ideas.
Have a vivid imagination.
Am not interested in abstract ideas.
Have excellent ideas.
Do not have a good imagination.
Am quick to understand things.
Use difficult words.
Spend time reflecting on things.
Am full of ideas.

C Arnett's Inventory for Sensation Seeking (AISS)

AISS describes two dimensions of sensation-seeking: novelty and intensity. *Novelty* measures the interest on exploring or discovering unknown things, places or people. *Intensity* measures the enthusiasm for intense sensations, such as gambling, horror/suspense movies, etc.

"For each item (20 in total), please indicate which response best applies to you: Describes me very well, Describes me a bit, Does not describe me very well, Does not describe me at all"

1. I can see how it would be interesting to marry someone from a foreign country.
2. When the water is very cold, I prefer not to swim even if it is a hot day.
3. If I have to wait in a long line, I'm usually patient about it.
4. When I listen to music, I like it to be loud.
5. When taking a trip, I think it is best to make as few plans as possible and just take it as it comes.
6. I stay away from movies that are said to be frightening or highly suspenseful.
7. I think it's fun and exciting to perform or speak in front of a group.
8. If I were to go to an amusement park, I would prefer to ride the rollercoaster or other fast rides.
9. I would like to travel to places that are strange and far away.
10. I would never like to gamble with money, even if I could afford it.
11. I would have enjoyed being one of the first explorers of an unknown land.
12. I like a movie where there are a lot of explosions and car chases.
13. I don't like extremely hot and spicy foods.
14. In general, I work better when I'm under pressure.
15. I often like to have the radio or TV on while I'm doing something else, such as reading or cleaning up.
16. It would be interesting to see a car accident happen.
17. I think it's best to order something familiar when eating in a restaurant.
18. I like the feeling of standing next to the edge on a high place and looking down.
19. If it were possible to visit another planet or the moon for free, I would be among the first in line to sign up.
20. I can see how it must be exciting to be in a battle during a war.

D Locus of control

1. Do you feel that wishing can make good things happen?
2. Are people nice to you no matter what you do?
3. Do you usually do badly in your schoolwork even when you try hard?
4. When a friend is angry with you is it hard to make that friend like you again?
5. Are you surprised when your teacher praises you for your work in school?
6. When bad things happen to you is it usually someone else's fault?
7. Is doing well in your schoolwork just a matter of "luck" for you?

8. Are you often blamed for things that just aren't your fault?

9. When you get into an argument or fight is it usually the other person's fault?

10. Do you think that preparing for things is a waste of time?

11. When nice things happen to you is it usually because of "luck"?

12. Does planning ahead make good things happen?

E   Closeness Measure

Table A1: Questions from friendship questionnaire used to construct the closeness measure

Question	Score
1. Do you know this friend's birthday?	1
2. How much time do you spend together in school, outside lessons?	
a) Most of the time	2
b) Some time	1
c) Hardly any time	0
3. How much time do you spend together in the school holidays? (Please mark one box only).	
a) I see this friend more than once a week	2
b) I see this friend about once a week	1
c) I see this friend less than once a week	0
4. What do you do together outside school? (You can mark more than one box).	
a) Shopping	1
b) Watching TV or listening to music	1
c) Going out at night	1
d) Playing computer/console games	1
e) Just talking	1
f) Other, please say what:	1
5. Do you talk to this friend about these things? (Please mark one box on each line). (if Yes=)	
a) Music	1
b) TV	1
c) Clothes	1
d) Films	1
e) Books and magazines	1
f) Sports	1
g) Politics	1
h) Computer games	1
6. Have you talked to this friend about a problem in the last week?	1
Maximum possible score (if friend and respondent in same school)	**20**