19,06,2012

UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS AND SOCIAL SCIENCES

Division of Economics, School of Social Sciences

Essays On Social Networks, Participation And Outcomes In Education

by

Greg Michal Bulczak

Thesis for the degree of Doctor of Philosophy

June 2012

ABSTRACT

This thesis explores the role of social networks in determining adolescents' outcomes in schools. The thesis consists of three papers that seek to empirically test how characteristic of friendship networks and peers affect adolescents' choices and performance in education. The main goal of the first paper is to estimate the effects of ego's friends age diversity on academic performance. The findings provide evidence that having an age diversified friendship network results in significantly worse academic outcomes. Contrary to the previous research, no evidence is found that having a best friend of a different age, or a group of friends of average age that differs from an individual's age is associated with worse outcomes in education. This paper addresses concerns about self-selection into networks and unobserved school level differences by using within-school variation and instrumental variable methods. The findings remain robust after the sample is limited to students with no criminal background and those that are in the expected grade for their given age. In the second paper a hypothesis that more interconnected networks (those with high density of friendships) positively impact on adolescents' school performance due to more scope for norms and sanctions, is tested. The findings provide evidence that for an individual having a close network during high school results in significantly better academic outcomes. Individuals with friends that know each other are found to be more likely to go to college. This examination addresses concerns about self-selection into networks and unobserved school level differences. Instrumental variable approach is used to investigate the effects of closure on college attendance. The effects of closure on years of schooling are found to persist for both low and high quality networks. The findings remain robust for samples consisting of non-white and white individuals. The last paper takes a closer look at participation in extracurricular activities, a factor that is likely to influence network formation. In this chapter, the role of community composition in determining participation outcomes is examined. This investigation provides evidence suggesting that racial composition of communities affects adolescents' participation in school extracurricular activities. The main contribution of this chapter is that problems related to sorting within communities and selection into schools, are carefully addressed.

Contents

List of tables

DECLARATION OF AUTHORSHIP

I, Greg Michal Bulczak

declare that the thesis entitled

"Essays On Social Networks, Participation And Outcomes In Education"

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission,

Signed: ……..

Date:…….

Acknowledgements

This research project would not have been possible without the support of many people. I wish to express my gratitude to:

- My supervisors, Jackie Wahba and Christian Schluter who were always helpful and offered invaluable assistance, support and guidance during my studies in Southampton;
- Mirco Tonin, Martin Gervais, Ronaldo Nazare, Chung-Yang Yeh and Juan Correa Allamand, for valuable ideas and comments.
- All members of the Economic Division for creating a professional working environment.
- The Economic and Social Research Council for providing the financial means.
- Russel Hathaway, Richard Udry and The Carolina Population Center for assistance with the Add Heath.
- All my graduate friends, especially officemates for sharing the literature and invaluable assistance.
- My family for understanding and support through the duration of my studies.

Note: This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

Chapter 1

Overall Introduction

The characteristic of individuals and the connections between them that form networks are likely to have measurable consequences for a number of reasons that will be discussed in detail in the chapters that follow. Social networks affect the flow and the quality of information, reward or punishment of members and trust. A growing body of research offers evidence that social networks play an important role in determining a wide set of outcomes. There are many examples of empirical investigations where networks influence business, employment or delinquency (Calvó-Armengol et al. 2009; Calvó-Armengol and Jackson 2004; Liu et al. 2011; Patacchini and Zenou 2008). Granovetter (2005) offers valuable insight about the effects of network structure on economic outcomes. Recent survey by Jackson (2010) discuses numerous studies in the area and explains the main problems associated with the analysis of social networks.

The central theme of this thesis, adolescents' performance and attainment in education, is closely connected to previous research that shows that various aspects of social structure impact on individuals' performance in schools. These aspects include differences in racial and gender composition of schools, grade cohorts and classrooms (Galeotti and Mueller 2005, Babcock 2006, Biffucclo 2011); and variations in grade span or the effects of shifting grades (Cook et al. 2008, Hattie 2002 and Hoxby 2000). It is argued that the influence of peers is likely to overcome the influence of parents or teachers in determining numerous outcomes (Steinberg et al. 1996).

Despite, the vast body of research there are still many uncertainties regarding how particular features of networks or different settings affect individuals' outcomes. These uncertainties arise mainly due to the fact that social networks are not normally observed. This results in difficulties with finding suitable measures of network structure and sources of exogenous variation in networks measures (Manski 2000; Jackson 2009). This thesis aims to overcome these difficulties with the use of a unique data set, the National Longitudinal Adolescent Health Survey (Add Health) where networks are directly observed and econometric techniques that allow identifying sources of exogenous variation in network structure. The uniqueness of the Add Health that comes from the extensive set of variables and directly observed friendship ties, allows conducting an analysis of previously unexplored areas of research. The key objective of this thesis is to investigate the role of social networks in determining adolescents' schooling outcomes. In other words, the goal is to examine the role of social capital in human capital formation.

The main contribution of this work to the existing literature comes from the analysis of the relationship between the characteristics of most direct friends and an individual's outcomes that can be linked to human capital accumulation. It is expected that this work will enhance knowledge about social structure and its impact on adolescents' attainment in education. The analysis that follows will provide valuable insight for designing schooling policies, such as grade span or size of schools, which may affect friendship formation in schools and similar setting.

The thesis consists of three papers (Chapters 2, 3 and 4) that seek to empirically test how various characteristics of friendship networks and peers affect adolescents' choices and performance in education. Chapters 2 and 3 look at the role of relatively close social networks that consist of best friends only. In these papers the relationship between friends' age, the density of friendship ties and performance in school is investigated. In Chapter 4, the social networks include not only best friends but also other peers from given grade cohorts in schools. Here, the impact of racial composition on participation in voluntary group activities is examined.

The main goal of Chapter 2 is to estimate the effects of friends' age diversity on academic performance. The findings provide evidence that having an age diversified

friendship network results in significantly worse academic outcomes. Furthermore, individuals in mixed-age friendship networks are found to be less likely to go to college. Contrary to the previous research, no evidence is found that having a best friend of a different age, or a group of friends of average age that differs from an individual's age is associated with worse outcomes in education. This chapter addresses concerns about self-selection into networks and unobserved school level differences by using within-school variation and instrumental variable methods. The findings remain robust after the sample is limited to students with no criminal background and those that are in the expected grade for their given age. Possible explanations for these findings include higher rates of conflict, less communication and cooperation in mixed-age friendship networks. The results find support in previous studies that linked ethnic and racial network diversity to lower levels of and adverse outcomes for individuals within them.

In Chapter 3 the analysis focuses on friendship ties, the fundamental element of social networks. A hypothesis that more interconnected ego networks (those with high density of friendships) positively impact on adolescents' school performance, is tested. The findings provide evidence that for an individual having a close network during high school results in significantly better academic outcomes. Individuals with friends that know each other are found to be more likely to go to college. This examination addresses concerns about self-selection into networks and unobserved school level differences. Instrumental variable approach is used to investigate the effects of closure on college attendance. The findings are in line with previous studies that find a significant relation between closure and outcomes. Furthermore, the analysis provides suggestive evidence that reputation is more likely to arise, and to have a bigger impact on attainment, in networks of high closure. The effects of closure on years of schooling are found to persist for both low and high quality networks. The findings remain robust for samples consisting of non-White and White individuals. Possible explanations for these findings include more pro-social behaviours associated with closure. The results find support in previous studies that link community closure and better outcomes for individuals.

Chapter 4 takes a closer look at participation, which is likely to influence network formation and to have a positive impact on other outcomes linked to human capital, development of social skills and social capital. In this chapter, the role of community

composition in determining participation outcomes is examined. This investigation provides evidence suggesting that racial composition of communities affects adolescents' participation in school extracurricular activities. The main contribution of this chapter is that problems related to sorting within communities, and selection into schools, are carefully addressed.

The three papers show that networks' structure plays a significant role in determining short and long-term outcomes for adolescents in high school. In all three cases a measurable relationship between different features of social networks and outcomes that can be linked to human capital accumulation has been identified. Results of the three chapters show that homogenous and highly interlinked networks improve an adolescent's outcomes in education. It is plausible that individuals in these networks benefit from better information flow and a greater scope for norms, trust and sanctions. This results in lower rates of conflict between members and increased satisfaction with schooling that are then reflected in better performance in education. Findings of this work contribute to the existing knowledge about the role of social structure in determining adolescents' decisions, choices and performance and may help to design better schooling environments in the future. This thesis opens many questions which due to time constraints and the design of the Add Health can be addressed in future analysis. This will be discussed in the chapters that follow.

The thesis is organised in the following way. The effects of age diversity on performance in schools are studied in Chapter 2. In Chapter 3 the role of network closure in determining adolescents' outcomes is examined. Chapter 4 focuses on the relationship between community composition and participation.

Chapter 2

Age Diversity Of Friendship Networks And Individuals' Academic Achievement.

2.1 Introduction

Peer networks, particularly when 'peer' refers to friendship, are one of the most influential factors affecting youths' behaviour, short and long-term academic, social and labour market choices. The impact of peer groups is often considered to be stronger than influence from parents or teachers (Haller and Ohlendorf 1970; Steinberg, Brown, and Dornbusch 1996). However, given the vast body of research focused on peer effects, the consequences for youth of having friends of various ages on academic performance remain unknown.

Peer effects research, its quantity and quality are often affected by the limited availability of data on social networks with many topics awaiting further investigation (Manski 2000). The Add Health data used in this study allows investigating various aspects of peer effects that remain largely unexplored. The first aspect arises from the fact that peer influences are frequently assumed to be uniform across age or grade mix, with previous research focused mostly on differences in schools' grade span or the effects of shifting six graders from elementary to middle school (Cook et al. 2008). Age diversity of friendship network is particularly important for adolescent, providing different sets of information, benchmarks for behaviour and group reward / punishment mechanisms. The second aspect is the lack or limited research investigating the diversity effect on long-term education choices including going to college. Instead most studies focused on peer influences in primary schools (grades 1-6) and test scores

(Hanushek et al. 2003; Henry and Rickman 2007; Kang 2007; Zimmerman 2003). Looking at high school students' outcomes is also more advantageous because there is more scope for age diversity in friendship networks due to the structure of secondary education in the United States, where mixed-age/grade classes are common. The abundance of elective classes and activities in high school gives more opportunity to interact and form friendships with students of different ages (Allan 1989). This gives more scope for age diversity to be important.

This research provides empirical evidence that age diversity of friendship networks matters for individuals' educational outcomes. There is a growing interest in the impact of group mix on various outcomes with most research focusing on racial or ethnic differences (Patacchini and Zenou 2006). Sandis and Hakuta (1999) offer a survey of studies examining educational outcomes and racial diversity. Other group characteristics such as age or skill differences that can influence workplace productivity or education have not attracted much attention largely due to the lack of suitable data.

Numerous studies confirm that social networks characteristics and one's position within them play an important role in various outcomes ranging from education, delinquency and criminal behaviour to job searching success (Jackson 2006; Calvó-Armengol, Patacchini and Zenou 2009; Ioannides and Soetevent 2006). Prior studies have also examined the effects of differences in racial and gender composition of schools, grade cohorts and classrooms on individuals' educational outcomes (Galeotti and Mueller 2005, Babcock 2006, Biffucclo 2011). In particular the negative effects of shifting six graders from primary to middle school have been proven (Cook et al. 2008). However, the effects of friendship network's characteristics on individual's performance were not captured by those studies and one can only speculate what changes happened to the individuals' friendship networks and the resulting consequences.

Given evidence from the studies mentioned above some questions arise: What is the impact of having friends of different ages on youths' performance at school? What is the effect of increasing one's network age diversity on his/her academic achievement? This investigation will focus on the effects of age diversity in adolescents' friendship network, as friends are a very special group of peers influencing individuals' behaviour and long-term choices.

In the light of limited direct evidence of network's age composition on performance to date, the central goal of this paper is to estimate the effects of age diversity in friendship network on educational achievement. Taking a closer look at peer group influence by investigating the peer effects of best friends, rather than on the composition of peers at school or grade cohorts level (which arguably forms only a background where friendship networks shape, giving scope for interaction and influence among individuals) will distinguish this research from previous body of literature. The main focus here is on examining whether students' grade point average and college attendance are influenced by age diversity of individual's social networks. Furthermore, this investigation will address concerns about endogeneity of networks' formation and unobserved school level differences by using within-school variation and instrumental variable methods (an individual's incidence of smoking). The intuition here is that smoking individuals have more scope for interactions with students of various ages; this provides opportunity for more mixed-age friendships with no direct impact on academic achievement, given controls. What further differentiates this investigation from previous studies on age diversity is that the primary interest is on the influence of closest friends, where other studies looked mainly at the immediate performance of structured groups (test or assignment groups) not allowing to pinpoint the influence of friendship networks or long-term schooling outcomes.

Examination of the impact of network's age diversity will provide valuable evidence for designing schooling policies that are likely to affect cross-ages friendship formation such as grade span, mixed-grade classes or organisation of extracurricular activities. In line with arguments presented by Allan (1989), more mixed-age friendships form in settings where there is an increased scope for interaction with individuals of different ages. In schooling systems with a short grade span or where students follow the same age cohort/class every school year, friendship networks are likely to be more homogenous with respect to age. It is expected that this research by examining the relationship between age diversity and attainment in schools will provide a new insight and contribute to the existing literature on the optimal schooling environment.

2.2 Literature review

2.2.1 Peer effects

The direct mechanism of peer effects points to the importance of peer interactions (Jackson 2010). Best friends constitute the innermost circle of students' social world in the school (Steinberg et al. 1996). Since students interact most frequently with their best friends, they are most influenced by these interactions. This influence is likely to be reflected in educational norms, expectations, motivation, learning habits and effort.

A number of studies suggests strong link between the academic achievement levels of college students' roommates and their own achievement levels (e.g. Sacerdote 2001 and Zimmerman 2003). Therefore, in studying the effects of peer group characteristics on students' educational outcomes, it is important to examine the effect of friends characteristics not only at the school or classroom level, where mostly indirect mechanisms of peer effects operate, but also at the individual friendship network level, where direct peer influence takes place.

Prior research looking at school performance focused mainly on three characteristics of peer groups: race, socio-economic composition and achievement; shows positive short-term effects of social origins desegregation on achievement and long-term effects (Hallinan 2001). Hattie (2002) compared multi-grade to single grade classes in primary schools and found no effect of class composition on outcomes. Analysis of 34 studies of primary schools conducted by Veenman (1995) found 21 of them indicating that being in a multi-grade school is associated with better attitudes toward school and self-concept. However, the general conclusion was that multi-grade classes being no worse than single-grade in terms of cognitive and achievement effects.

The problem of selection bias is likely to influence the above findings. Mason and Burns (1995) argue that multi-grade classes usually consist of better performing students and higher teaching quality; this masks the negative effect of multi age classes. Veenman (1996) re-examined the matter using meta-analytic approach and again found similar results indicating no difference in education achievement with respect to mixing

for grades 4-5, positive effect for grades 1-3. However, a negative effect for grades 6-7 was identified in the study. This is of particular interest, assuming that class composition can be treated as a proxy for friendship formation, with multi-grade classes creating more opportunities for heterogeneous age friendships.

However, these arguments apply to primary schools only, where multi-grade classes are often introduced out of necessity in order to cope with low number of students in a given year (most frequent in small schools). Individuals in primary schools are very unlikely to form mixed-age friendships (Cotterell 2007). Secondary education differs significantly for a number of reasons; more mixed-age friendships form, mixed-grade classes are common, with the number of elective modules (usually mixed classes) offered depending on schools' resources.

2.2.2 Diversity and performance of groups

Studies in the field of labour and organisational economics that investigated the impact of various dimensions of teamwork composition on performance have found inconclusive evidence (Reagans and Zuckerman 2001; Riordan and Shore L 1997). However, a more recent examination indicates that team diversity can be beneficial for productivity (Richard and Shelor 2002). Generally, more diversity is associated with a more diverse scope of knowledge and perspectives resulting in greater "diversity capital" (Arcidiacono and Vigdor 2010). On the other hand, counterproductive factors have also been identified including higher communication cost, lower trust and less interaction (Zenger and Lawrence 1989).

Mannix and Neale (2005) provide a summary of main findings in psychology. They conclude that general social category differences (race, gender, or age) are more likely to negatively influence the groups' performance. However, they also argue that differences in functional background, education, or personality, are more likely to have a positive impact on outcomes. The authors speculate that this positive effect arises because of increased creativity or better group problem solving ability. The study indicates that most of these effects can be linked to social integration, communication, and conflict.

The above findings relate to adults performance in teamwork settings and provide only limited insight about adolescents' friendship networks and academic performance. It is highly probable that the effects of group diversity are different for youths as opposed to adults and also to depend on the setting (work, primary or secondary school) or the type of groups. For example, Mason and Burns (1996) investigated students' productivity in solving mathematical problems in single and mixed-grade classes in primary schools and found that the latter had significantly lower cooperation and assistance levels. No evidence was found that mixing ages enables positive effect through peer tutoring or social development. More generally, heterogeneity in groups was found to affect members by decreasing satisfaction with the group, lower levels of cohesiveness, reduced within-group communication, decreased co-operation, increased turnover and higher levels of conflict (Triandis et al., 1993; Riordan and Shore, 1997). Prior examinations link community heterogeneity to reduced propensity to participate in social activities including education (Alesina and La Ferrara 2000). Mobius and Szeidl (2009) provide plausible explanation as to why diversity may be linked to worse outcomes. They construct a model in which the average trust in a network is monotonically decreasing in heterogeneity of a network (for simplicity they consider race and ethnicity, however, for adolescents' age is likely to be an equally strong factor). The logic is that networks with greater trust allow for more social sanctions between individuals, increasing incentives for cooperation.

2.2.3 Age Diversity and groups

Sociology and psychology offers interesting insight into some of the consequences of having different ages in friendship networks. Cotterell (2007) provides a review of previous social science studies examining the effects of age diversity on adolescents. Positive influences include more opportunities to develop personality and social competence, more diversified sources of information referring to activities and academic choices. The list of negative influences is much longer and includes: increasing number of models of bad behaviour, restricting network members from participating in various forms of school activity (considered unpopular or "not cool"), older friends are more likely introduce to drinking or smoking and to reduce importance of own classmates as a reference group (older peers presence serves as

example/benchmark of behaviour) resulting in accelerated exposure to adult lifestyle. Furthermore, it is emphasised that any deviation from age homogeneity is perceived negatively by adolescents, affecting adversely social status of an individual. This could result in increased isolation of a given individual and his network, consequently decreasing information flow to this individual.

Allan (1989) finds that as opportunities for cross age friendships increase, more mixed-age friendships form, with the total number of friendships staying the same and that social competence was highest for those with peers of the same age. Young people that have formed friendships with an older or younger friend, in consequence, had lower estimates of their own competence and popularity. Strong normative preference for same age friendship can be seen. It is suggested that heterogeneous networks do not fit the social norm, leading to social exclusion. The adverse effect of having mixed-age peers network is that having older or younger friends could be perceived as a bad thing for one's social standing within school environment. Students in mixed-age networks may suffer from network and individual isolation mechanisms affecting negatively the flow and quality of information. In consequence, less informed and demotivated students should be more likely to drop out of school and not to go to college. A dissonance exists as to whether having older friends helps to develop social maturity faster, with most studies not finding evidence that such relationship exists. Mason and Burns (1996) find that older peers provide a frame of reference that leads to devaluated competence/motivation and consequently less effort in education, and results in being more critical about themselves.

Generally, in the light of previous research, it is difficult to predict what will be the effects of age diversity on adolescents' attainment in schools. It is possible that students with friends of mixed ages will raise the cost of maintaining friendships and increase cognitive biasing, leading to less open communication and more conflict within networks. However, the literature also indicates other influences that may have a positive impact on some schooling-related outcomes.

2.3 The data and descriptive statistics

This study uses data from the public use version of the National Longitudinal Adolescent Health Survey (Add Health), which is a nationally representative sample of youths in secondary education. The Add Health data set was designed and conducted by the Carolina Population Center for the purpose of measuring the influence of social environment on adolescent health. The survey contains detailed information on general health and well-being of adolescents in the United States, including: the behaviours that benefit and influence health. Some of the dependent variables include diet and nutrition, violent behavior, intentional injury and suicide, exercise. The design of this survey and an extensive set of questions that cover various aspects of everyday behaviour of adolescents, allow conducting comprehensive investigation of peer effect, social networks and academic performance (To correct for the design effects of the survey, the Public-Use dataset, contains weight variables. These sampling weights provided in the Add Health are used in the analysis in this and the following chapters). The survey is based on 144 randomly selected schools in the United States in three waves 1994 - 1995 - 2002. Due to a low response rate in 15 schools, the data set provides information on 129 of the 144 total schools (6344 students). Complete network data was collected in every school at Wave 1 (1994). Each student was asked to name up to 5 male and 5 female friends. These data offer the basic information needed for network context measures. In addition to the nominations that students give, they also receive nominations from other individuals. The key benefit of received nominations is that they allow limiting the problem of the constraint imposed on individuals by only permitting to nominate five best male and five best females (send network constraint). From the initial sample of Wave 1 a sub-sample of individuals was selected to complete more detailed in-home questionnaire. The in-home sub-sample was interviewed again in 1995 (Wave 2 of the study) and 2002 (Wave 3). Wave 3 contains attainment variables that are of particular focus and will be used as dependent variables in Chapters 2, 3 and 4. They include final year grade point average (GPA), years of education completed and information on college attendance.

Friendships nominations data provide the basic information needed for network context measures such as age diversity of friendship networks. The diversity measure of networks, already provided in the dataset, assess the diversity of an ego network with respect to the traits of a categorical attribute (here age). The formula used to calculate ego-network diversity with respect to attribute A for a given individual, i , is as follows:

$$D_{iA} = 1 - \left[\sum_{1}^{n} \left(\frac{A_k}{en} \right)^2 \right],$$

where:
A = the attribute (e.g. age or race)
A_k = the number of nodes with trait *k* in the ego network
(e.g. individuals age=16 or white)
en = the number of nodes in the ego network with valid data on A
n = the total number of traits (ages) of A represented in the ego network
(e.g *n*=1 if network consists of same age or same race individuals.

The dataset allows distinguishing between tree types of ego networks: send, receive and send and receive. This study will focus mainly on send and receive and send (out – degree networks). The diversity measure based only on sent nominations allows assessing the most direct impact of friends that an individual (ego) considers to be friends. The inclusion of received nominations allows for more variation, also there is no constraint on the number of received nominations, so send and receive networks can be noticeably larger. Also the benefit of including received nominations is that individuals that are not necessarily most direct friends are accounted for. In this, and the following chapters the term network will be used with reference to an ego's send and receive friendship network. In the cases where the analysis will look at send or receive networks, this will be clearly stated.

Table 2.1 provides descriptive statistics for selected variables that are used in this chapter. The variables include the measure of main interest - network's age diversity and age-related indicators. It can be observed that most of the students have friendship networks that are not highly diversified with respect to age. 27% of the adolescents reported having a significantly (more than one year) older best friend. Furthermore, in 22% of cases the difference between the age of an individual and the average age

his/her friendship networks was more than a year. In addition, the table contains information about an individual's incidence of smoking, which will be later used to instrument for networks' age diversity. It appears that approximately half of the sample consists of non-smokers.

Table 2.1: *Summary statistics.*

Variable	Minimum	Maximum	Mean	Std.
Age Diversity	0.00	0.82	0.26	0.23
Age Diversity (sent)	0.00	0.81	0.27	0.26
GPA Final Year	0.14	4.00	2.83	0.82
College	0.00	1.00	0.39	0.48
Years of Education	0.00	15.00	8.97	6.15
Incidence of Smoking	0.00	30.00	8.43	7.13
Smokedin30days	0.00	1.00	0.29	0.45
Smoked never	0.00	1.00	0.52	0.49
Older best friend	0.00	1.00	0.27	0.44
Younger best friend	0.00	1.00	0.23	0.42
Best friend of different age	0.00	1.00	0.46	0.50
Younger network	0.00	1.00	0.14	0.35
Older network	0.00	1.00	0.07	0.26
Avg. netw. age≠ ind's age	0.00	1.00	0.22	0.41

Note: Number of observations: 4764.

Table 2.2 provides means for selected variables for high and low age diversity levels (the split divides the population into two approximately equal parts). Students in highly age-diversified networks have lower ability test scores and GPA. Those in networks with low diversity are more likely to go to college (68% compared to 55% for those in high diversity). Individuals in mixed-age friendships are also more likely to smoke and have mothers with lower education levels. Race, age or gender differences between the two groups are not significant. Finally, the levels of criminal behaviour /gang membership proxied by stealing behaviour are not different.

Table 2.2: *Comparison by high and low age diversity groups.*

	Network diversity	
variable	high mean	low mean
College	0.55*	0.68*
GPA	2.67	2.85
Years of Education	7.12	10.03
Peabody test	50.75*	59.31*
Male	0.49	0.48
Non-white	0.34	0.33
Age	15.89	15 .67
Tot. network centrality	0.83	0.71
Incidence of smoking	10.67*	8.22*
Drugs user	0.13	0.13
Drinks alcohol	0.36	0.33
Steal	0.28	0.29
Mother ed low	0.15*	0.10*
Mother ed med	0.31	0.29
HH income	1196.09	1107.11
Urban area	0.53	0.55

Note: *- indicates a significant difference low vs. high diversity at the 95% confidence level. The split divides the population into two approximately equal parts (46% of the sample in the low category.

2.4 Empirical analysis

This study will examine the relationship between age diversity and academic achievement outcome measures including: college attendance and GPA. Each of these variables is reported in Wave 3 (year 2002) of the study. GPA refers to the average of all grades based on transcripts from high schools in the final year. The literature discussed in the previous section does not directly indicate what will be the impact of age diversity on an individual's academic performance.

The aim is to determine what role age diversity of an individual's network has in determining academic outcomes of students. More formally, the reduced form model is:

(1) $Y_{ij} = \beta_1 D_{ij} + \beta_2 X_{ij} + \beta_3 S_j + \varepsilon_{ij}$

Y_{ij} is an education outcome measure for individual i from school j observed in wave 3,
D_{ij} is heterogeneity measure of individual's friendship network,
X_{ij} is a set of control variables,
S_j school dummies,
ε_{ij} is the error term.

A set of control variables in eq.1 is used to limit the possibility that the relationship between age diversity and academic outcomes could arise because of correlations between academic performance and characteristics related to individual specific factors (students living in low income neighbourhoods, living with poorly educated parents, propensity to take risk proxied by recent injury and reported seatbelt use, individuals of low ability or with criminal background/gang membership). In addition, variables related to the quality of friendship network will be included (Network's GPA and delinquency). A related measure, total friendship network centrality (position within the whole school network), will be also added. It is expected that this measure will act as a proxy for an individual's social skills and position within the school network. It is quite likely that individuals on the outskirts of the total school network will be very different from those centrally located. In the light of studies discussed in the previous section, this variable may have a significant impact on educational outcomes as well as on age diversity. Furthermore, exposure to similar school environments that are not observed

may be a problem. School characteristics may impact on peer influence and performance (contextual effects may include racial composition, poverty levels and parental education). The Add Health study interviews all adolescents within a given school allowing estimation conditional on school specific effects. The inclusion of school specific factors (S_j) means that only the variation in age diversity of individuals' friendship networks across individuals in the same school is exploited. It is also important to control for neighbourhood quality as it is possible that deprived neighbourhoods will be more likely to provide increased opportunities for mixed-age friendships (less parental supervision or lower number of organised activities). Median neighbourhood income will be used as a proxy for the quality of the area where individuals live.

Ordinary least squares (OLS) estimates are presented in Table 2.3. The coefficient on age diversity of an individual's friendship network is negative and statistically significant. The results also indicate that the quality of friends measured in terms of GPA has a positive impact on final year GPA. Generally, the coefficients on the remaining variables are in line with previous studies (e.g. Allcott et al. 2007; Rees and Sabia 2010).

Table 2.3: *Age diversity and academic attainment.*

Outcome:	Final GPA		College		YOE	
	Coef.	P>|t|	Coef.	P>|z|	Coef.	P>|z|
Age Diversity	-0.297	0.002	0.364	0.005	-0.324	0.085
Friends' GPA	0.299	0.000	2.073	0.000	0.721	0.000
Network Centrality	0.090	0.012	1.664	0.001	0.595	0.028
Peabody test	0.005	0.000	1.009	0.000	0.007	0.000
Male	-0.322	0.000	0.608	0.000	-0.224	0.000
Non-white	-0.029	0.571	1.437	0.042	0.458	0.000
HH income	0.001	0.013	1.006	0.182	0.001	0.001
Parental involvement	0.027	0.001	1.067	0.025	0.097	0.000
Mother edu. high	0.043	0.573	3.283	0.000	1.032	0.000
Mother edu. med	-0.028	0.718	1.855	0.009	0.711	0.009
Steal	-0.038	0.334	0.958	0.761	-0.380	0.782
Delinquency network	-0.104	0.021	0.690	0.015	0.004	0.972
Expected age	0.191	0.028	2.252	0.008	0.192	0.048
Health	-0.033	0.112	0.818	0.004	-0.326	0.377
Seatbelt wearing	-0.002	0.889	1.053	0.378	0.473	0.121
Injury	-0.011	0.629	1.044	0.529	0.043	0.278
Urban	-0.055	0.357	1.104	0.644	0.010	0.923
School dummies	yes		yes		yes	
R squared	0.42		0.30		0.43	

Note: N. of obs.: 4764. In column 2 exponents of the logistic regression coefficients are reported. Additional controls include age, drugs and alcohol consumption, attractiveness of personality assessed by the interviewer, propensity to take risk proxied by seatbelt use, variables controlling for of siblings' presence and age, indicator of membership in clubs/societies, grade cohort dummies, controls for network composition (density and heterogeneity with respect to race).

In column 2 of the table the outcome measure (GPA in the final year) is replaced by a college attendance indicator (equal to 1 if the individual continued education after high school). The impact of diversity is also negative and significant; suggesting that having friends of mixed-ages in high school reduces the probability of attending college in the future. Most of the coefficients reported in the table have the expected sign and significance in line with the results reported in the first column where GPA in the final year was the outcome of interest. Differences in the level of significance can be observed for mother's education and Non-white variables. In column 2 both are found to significantly increase the chances of going to college. In both cases, GPA and college attendance, it is possible to control for previous school performance (GPA). However, GPA and Peabody test scores are highly correlated and when both included in the regressions, the results misleadingly indicate weakly significant role of ability in

determining outcomes in education. For this reason only the Peabody measure is included in the regressions. This does not significantly affect other coefficients. In addition, column 3 the outcome measure is years of education completed by an individual. Here, the results are in line with estimates presented in column 1. It can be noticed that the significance of the key interest drops.

The above results provide suggestive evidence that diversity impacts negatively on adolescents' performance in schools. However, even with the rich set of controls it is still possible that due to problems that will be discussed below the estimates do not capture the true impact of age diversity in friendship networks. Next, possible problems and ways to alleviate them are discussed.

2.4.1 Selection problem

Manski, in a series of key articles (1993 and 2000), has underscored many problems of interpreting outcomes that result from the estimation of social interactions. A more recent study by Bramoulle et al. (2009) describes the main obstacles and ways to tackle them. For this study the key issue in identifying network effects is that individuals or households often have some scope for choice of peer groups, whether through the selection of place of residence, school, or friends. Given that this investigation aims to pinpoint the effects of an individual's network age diversity on education outcomes, it is possible that after controlling for selection across schools, the coefficient measuring the impact of diversity, β_1 in eq.1, is still biased due to self-selection into friendship networks. This may arise if there are unobservables that affect outcomes in education and age diversity of friendship networks that are not included in equation 1. One way to tackle this problem is to use the instrumental variable (IV) method. A valid instrument would make age variation in an individual's friendship network that is "as if" randomly assigned. The following part of this investigation is primarily focused on identifying a valid instrument.

Equation 2 is the reduced form model of age diversity aiming to identify factors that determine age diversity of an individual's network.

(2) $D_{ij} = \beta_1 F_{ij} + \beta_2 S_j + e_{ij}$

D_{ij} is a network diversity outcome measure for individual i from school j,

F_{ij} is a set of variables, including incidence of smoking, likely to affect age diversity of an individual's friendship network, (assumed to have no direct influence on Y_{ij} ; the variables are listed in Table 2.4)

S_j stands for school dummies

Coming back to the problem of selection on unobservables into friendship networks, it is possible that *e* in eq.2 and ε in eq.1 are correlated due to unobservable variables that influence outcomes in eq.1 and 2. If this is the case, β_1 in eq.1 is biased. One may speculate that more socially mature individuals are more capable of maintaining mixed-age friendships and they also may be more likely to maintain good relationships with teachers, plan ahead and do better at school (the IQ proxy used as a control may not be a proper one for social ability). Furthermore, better performing students may be more likely to help (at own initiative or asked by teachers) struggling students, resulting in more opportunities to form friendships with students of different ages. Not accounting for this may result in underestimation of β_1 in eq.1. On the other hand, factors that inflate the coefficient are arguably easier to control for by including indicators of repeating a year and proxies for criminal propensity/gang membership.

Equation 2 is used to identify determinants of age diversity and to test for the relevance of instrumental variable candidates. Of particular focus is an individual's smoking behaviour. The intuition is that there exist differences in cigarettes availability for different age groups (18 is the legal age to buy cigarettes in the United States). So for those in the final years of high schools (closer to the age requirement) obtaining cigarettes should be easier. Younger students facing more difficulty in buying cigarettes could use older friends as intermediaries. Cotterell (2007) presents arguments that further support the above intuition. Smoking is a social activity that is likely to be

performed in hiding or secret places when at school, giving scope for further interaction between students of different ages when smoking during breaks between classes or before/after school day. As a result, smokers should have more mixed-age friendships. The first-stage results (Table 2.4) indicate that the incidence of smoking measure meets the criteria for a relevant instrument (p-value is less than 0.01).

Table 2.4: *Reduced form model of age diversity. First stage regression.*

Diversity	Coef.	p-value
Smoking incidence	0.004	0.001
Drinks alcohol	-0.008	0.454
Drugs user	0.023	0.147
Expected age	-0.043	0.042
Age	0.003	0.021
Male	-0.014	0.154
Steal	0.006	0.128
Peabody test	0.001	0.024
Netw. Centrality	0.048	0.000
Mother edu. high	-0.006	0.545
Urban area	-0.004	0.807
School dummies	yes	
R squared	0.21	

Note: *N. of obs. :* 4764. *Additional controls as in Table 2.3*

For smoking to be a valid instrument it must not only be relevant but also exogenous. The exogeneity assumption requires that after controls, smoking has no effect on outcomes of interest (GPA or college attendance). Smoking cigarettes arguably has no direct effect on GPA. No empirical evidence exists that directly links smoking to academic performance. Smoking does not result in such adverse consequences as drinking alcohol, which may affect students' performance at school, particularly the study time, test taking and ability to concentrate in class (Wolaver 2008). A study by Burt and Peterson (1998) finds that school smoking cessation behaviour has no association with students' GPA. Unfortunately, in the case of one instrument, exogeneity cannot be tested. Therefore, it is particularly important to carefully address possible problems associated with using incidence of smoking as an instrumental variable. A number of previous studies point out important factors that need to be controlled for (Cook and Hutchinson 2007, Eysenck 1991). One aspect that calls for attention is personality. Previous research suggests that particular personality characteristics (being extrovert) may affect academic attainment (Kézdi and Cseres-

Gergely 2008). In the case of extroverts a study by Canals et al. (1997) finds no significant link between this particular trait and adolescents smoking. Kubicka et al. (2001) also find no evidence that would associate personality characteristics with smoking behaviour. The only characteristic that appears to matter is risk taking as suggested by Burt et al. (2002). If risk taking also affects grades or college outcomes, and is not controlled for, then the incidence of smoking would not be a valid instrument. To eliminate this possibility individuals' risk taking will be controlled for. Two measures will proxy individuals' risk taking: wearing a seatbelt while in a car and body injuries in the past year. Hersch and Pickton (1995) find evidence that seat belt users are more risk averse. Body injuries are likely to be good indicators of past risky behaviours. In addition, delinquency, health and peer quality controls will be included in the analysis as they may be linked with an individual's risky behaviour. Furthermore, the variables already included in the previous specification control for the possible influence of commitments, social bonding and peer group structure on smoking behaviour (Ennet and Bauman 1993; Krohn et al. 1983).

When incidence of smoking is included in the OLS regression where GPA in the final year is the outcome of interest; the coefficient is very small and statistically insignificant (coef. = 0.0006, p-value= 0.67). This offers only suggestive evidence in support of the above arguments. Assuming that the instrument meets the exogeneity requirement, it can be plausibly excluded from the outcome regression. Therefore, using the incidence of smoking as an instrumental variable limits the possibility of obtaining biased results by making age diversity as if randomly assigned.

The results of Table 2.4 also show that gender or delinquency do not influence the diversity measure. In line with previous expectations is the significant and positive influence of an individual's age. The Peabody vocabulary test proxying for individuals' intelligence is also identified as a significant positive factor; however, the size of the coefficient indicates that the influence is relatively small.

The next aim is to estimate the effects of network age diversity on academic outcomes with the help of previously introduced instrument. Equation 3 is the second stage outcome where β_1 measures the effect a friendship network's age diversity:

(3) $Y_{ij} = \beta_1 D^*_{ij} + \beta_2 X_{ij} + \beta_3 S_j + \varepsilon_{ij}$

Y_{ij} is an education outcome measure for individual i from school s,

D^*_{ij} is the age diversity measure of individual's friendship network instrumented by the Incidence of smoking variable,

X_{ij} is a set of control variables

S_j school dummies

ε_{ij} is the error term

Table 2.5 presents results from the second stage outcome regression. It can be observed that the coefficient of interest drops both in terms of size and significance when compared to the OLS estimates. Most likely the effect of age diversity was overestimated due to the problems discussed above. Generally the results are in line with previous research aiming to identify determinants of GPA. They indicate that having an age diversified friendship network has adverse effect on students GPA in high school (1% increase in the diversity measure lowers GPA in the final year by .11). This could be due to previously discussed problems that diversity in networks is associated with; including less trust and communication, more conflict, higher cost of maintaining friendships and less satisfaction with schooling. The coefficients on variables indicating gender and ability have the expected signs and are statistically significant. Those students that admitted drinking alcohol are found to be more likely to have lower grades. Another result consistent with previous findings (Calvó-Armengol et al. 2005) is that individuals located more centrally in their school's friendship network are more likely to do well.

Table 2.5: *The second stage regression results.*

Outcome	IV (smoking)		Non-criminals		Right - age sample	
GPA final year	Coef.	p-value	Coef.	p-value	Coef.	p-value
Age Diversity	-0.109	0.047	-0.125	0.051	-0.112	0.043
1 stage p-value		0.001		0.001		0.001
Additional controls		yes		yes		yes
School dummies	yes		yes		yes	
R^2=	0.34		0.33		0.33	

Note: N of obs.: 4764, the sample is reduced by 1832 observations in the second column and by 737 in column 3. The regression includes controls as in Table 2.3.

Due to the worries that the results may be partially driven by students that are not in the expected grade for their age (10% of the sample) the estimation is repeated on a restricted sample where those individuals are excluded. Of particular concern was that those students may be of significantly different than the expected age for their given grade due to differences in academic performance or health problems in the past. It follows that those not in the expected grade would more likely have more diversified networks with respect to age. This exclusion does not change the results of the analysis (right - age sample, reported in Table 2.5) significantly. Similar steps with similar results (Table 2.5) are taken with respect to individuals that admitted stealing private property in the past (30% of the sample). After limiting the sample to those with no criminal background the significance of most of the factors decreases slightly. This can be partially due to the large reduction in the sample size.

2.4.2 Additional Results

Next an alternative age diversity measure is included in the analysis. This measure is based on out-degree nominations, where only friendship nominations sent by ego count. The key difference between this measure and the previously used one is that for an individual to be included in ego's friendship network it is necessary to be named as a friend by ego. This difference in the constructions allows looking at networks that are arguably closer to ego and may be more likely to impact on his or her outcomes. The results presented in Table 2.6 show that the coefficient of main interest increases in size, however its significance decreases. The increase in size may be explained by the greater role of closer friends in affecting the outcome of interest. Lower significance may come from the exclusion of unwanted friends (those that nominated ego with no reciprocity). It is possible that age diversity amongst those friends has more significant impact on individuals' performance in education.

Table 2.6: Additional results: *Out-degree and high and low diversity.*

Outcome	IV (smoking)		High diversity		Low diversity	
GPA final year	Coef.	p-value	Coef.	p-value	Coef.	p-value
Age Diversity (sent)	-0.137	0.101				
Age Diversity			-.181	0.044	-0.073	0.008
1 stage p-value		0.001				
Additional controls		yes		yes		yes
School dummies	yes		yes		yes	
R^2=	0.33		0.33		0.33	

Note: N of obs.: 4764, in columns 2 and 3 the sample is divided into approximately to equal part. The regression includes controls as in Table 2.3.

The table also presents results split by low and high age diversity. In order to test for possible asymmetric effects the sample is split into two. The split is made for low and high network age diversity (it divides the sample into approximately two equal parts). The results are presented in columns 2 and 3 of the table. It appears that the changes in diversity levels have larger effect on GPA in the high diversity networks sub sample. However, the negative effect remains statistically significant in both cases. This finding has support in the previously discussed arguments that high levels of diversity may have the most adverse effects on performance.

In addition, tests were performed to examine whether the relationship between age diversity and achievement may be non-linear. For example, it may be the case that moderate levels of diversity have a positive impact and that very low and very high levels affect schooling outcomes negatively. In this case the relationship would be a U-shape relation. To test that hypothesis a squared term of age diversity is introduced in the regression. The coefficient on the term is not statistically significant (not reported here), suggesting that the U-shaped relationship is not likely.

Next part of this analysis aims to assess the effects of having an older best friend or a significantly higher/lower average age of friends. Because the diversity measure is not sensitive to the size of age gap between an individual and his/her friendship network it is of interest to see if the gap has a measurable effect on outcomes of interest. The aim of the next exercise is to test whether in addition to the negative effects of diversity, having an older or younger friend affects academic performance. First, binary variables indicating a best friend that is older/younger than the individual (equal to 0 if same age) are added to the regression. The intuition is that age of the best friend may have a special role in determining an individual's behaviour. An older friend may be a mentor, providing guidance and information about which classes to take in the future, teachers that one should avoid or college enrolment. On the other hand, as previously mentioned, a younger best friend may also be of benefit by facilitating the development of social skills linked to mentoring. Results reported in Table 2.6 show that having a younger or older best friend is weakly associated with higher GPA in the final year. Of particular interest is that the significance of age diversity remains unaffected after best friend dummies are added. This allows eliminating the possibility that age diversity is simply proxying for the best friend's influence. What is more, it appears that the measures of age diversity and best friend's age have the opposite effects on school performance. This finding has some support in the literature related to mentoring and personality development of adolescents (Cotterell 2007). In particular, it is argued that having a younger best friend provides opportunities to develop skills linked to mentoring and ability to guide others by setting an example. On the other hand, an older best friend may play a positive role by providing information about future academic opportunities, setting an example and giving guidance.

Table 2.7: *The effect of having a friend or friends of different age.*

GPA final year	Coef.	p-val.	Coef.	p-val.	Coef.	p-val.	Coef.	p-val.	Coef.	p-val.
Age Diversity (instr.)	-0.099	0.048	-0.101	0.048	-0.094	0.049	-0.103	0.010	-0.044	0.021
Older best friend	0.180	0.080								
Younger best friend			0.270	0.130						
Best friend ≠age					0.320	0.060				
Avg.net.age≠ind. age							0.414	0.038		
Younger network									0.291	0.044
Older network									0.405	0.121
School dummies	yes		yes		yes		yes		yes	
R^2=	0.370		0.370		0.370		0.370		0.370	

Note: N of obs.: 4764, the regressions include controls as in Table 2.3.

To summarise, after accounting for a network's age diversity, different age of best friend has a weak positive impact on an individual's performance. When the estimation is repeated with the diversity measure missing, the results misleadingly indicate that a best friend of a different age has a negative impact on performance. The above exercise shows that the negative effect is not due to the best friend's age but the diversity of the total friendship network. It would be of interest to exclude the best friend from the age diversity measure, however because of the nature of the public use data set this is not possible.

Concentrating on the best friend provides interesting insight into the role of the closest peers but tells very little about the rest of the friendship network because of the restricted scope for social interactions. For this reason another measure, the average age of friendship network, is used to check if it may provide more intuition. Again, a variable indicating networks' average age significantly (more than 12 months) lower or higher than the individual's age is entered in the regression. From results reported in Table 2.6 it appears that an individual of age that is significantly different than the average age of his friends performs better in terms of GPA. A further test, where two variables indicating networks' average age significantly lower or higher than the individual's age are entered in the regression, shows that this positive effect is mostly due to the lower average age. As to why individuals in younger networks (in other words individuals older than their networks' average age) perform better is an open question.

2.5 Discussion

Generally the results show that age diversity of friendship networks adversely affects academic performance of adolescents. The negative effect of having friends of various ages is not only specific to the GPA measure but also has a significant impact on decisions whether to attend college. Furthermore, having a best friend of a different age, or a group of friends of average age that differs from an individual's age, is not associated with worse outcomes in education. The findings indicate that the opposite is more likely to be the case. In the light of previous diversity - related research on adults in workplace settings and controlled groups, it seems plausible that similar explanations for the negative effects of diversity can be applied in the case of adolescents and friendship networks. The most likely channels trough which diversity can be linked to poorer performance include lower levels of trust in heterogeneous networks (Mobius and Szeidl 2007), and higher rates of conflict, less communication and cooperation in mixed-age settings (Zenger and Lawrence 1989). It is highly probable that similar forces are at work in a controlled group setting and in friendship networks. Due to those factors, linked to age diversity, adolescents with friends of various ages could experience less backing from their peers and find schooling experience less rewarding and for that reason achieve less in school.

The results find further support in previous research connecting community heterogeneity to reduced propensity to participate in social activities including education (Alesina and La Ferrara 2000). This is in line with this study's finding linking age diversity to worse performance in schools. Additional support comes from Mobius and Szeidl (2007) where the average trust in a network is monotonically decreasing in heterogeneity. Likely explanations for bad outcomes of individuals in mixed-age networks in this study are lower levels of trust and more conflict in these networks. Results by Glaeser et al. (2000) are supportive of this interpretation suggesting that this is most likely to affect negatively the benefits students obtain in schools, having a direct impact on GPA and college attainment.

It is important to emphasise the limitations that arise due to the nature of the data and the empirical strategy of this investigation. They arise mainly because the public use version of the Add Health data set does not provide adequate information about an individual's reference group (for example friends' identification numbers are not accessible). In consequence, the specification of the reference group is a result of the data set design rather than desired criteria. One limitation of this study is that the "reflection problem" as described in Manski (2003) does not allow for clear indications for policy making. The models presented in this analysis do not account for the Spatial Autoregressive (SAR) nature of the data. If there is too much dependence in the data the resulting estimates may be biased and inconsistent. A possible solution would be to use a SAR model, which could bring additional source of information and could be valuable for identification or improving estimation efficiency as in Lin (2009). In SAR models the "reflection problem" is eliminated by relying on the non-linearity introduced by the variations in the peer measurements.

Furthermore, using smoking behaviour as an instrumental variable is associated with significant obstacles. Particularly, in the light of previous studies it appears that it cannot be used for college or years of education outcomes. The use of smoking behaviour in order to instrument for age diversity, when final year GPA is the outcome of interest, is also challenging because one needs to very carefully control for an individual's risk taking and social attitudes. Ideally, to relay on the IV estimation, a second instrument would be needed. In this case the exogeneity assumption could be empirically tested.

In addition, the narrow set of outcomes that is offered in the public use data set is constraining. The negative relationship between age diversity and attainment could be easily used as an argument for single grade classes. However, it is likely that there are numerous benefits of having friends of various ages such as greater ability of working with diverse teams or development of mentoring skills that are not investigated in this study. One possible direction for future study is to examine how the characteristics of social networks studied here impact on other outcomes such as employment prospects and wages.

Another area that this chapter leaves unexplored, due to the data limitations (the public use version of the Add Health does not permit constructing new network variables), is how robust are the results to other measures of diversity. Ideally, an alternative measure for age diversity would be sensitive to the size of age differences between a network's members.

2.6 Conclusions

Results of this study indicate that age diversity of an individual's friendship network has negative influence on future academic achievement. Having friends of different ages in a network is associated with lower GPA in the final year of high school and a lower probability of having attended college. The results are in line with previous studies that find a significant relation between community diversity and poorer outcomes for individuals, suggesting lower levels of trust as the main explanation for weaker academic performance of individuals in diverse networks. Contrary to the previous research no evidence is found that having a best friend of a different age, or a group of friends of average age that differs from an individual's age is associated with worse outcomes in education. The results remain robust after the sample is limited to students with no criminal background and those that are in the expected grade for their given age. It is important to emphasise that age diversity may have a positive impact on other outcomes not directly linked to outcomes considered in this analysis such as the development of various social skills.

With previous studies investigating the effect of diversity on performance in much more controlled settings, such as projects assignments or firms, the primary contribution of the paper is measuring the effects of network age diversity on individuals' performance in an adolescent's friendship network setting. The clear benefit of looking at best friends rather than at a school or grade level is that the most direct peer influences are observed. Taking a closer look at peer group effects by investigating the peer effects of best friends, rather than on the composition of peers at school or grade cohorts level (which

arguably forms only a background where friendship networks shape, giving scope for interaction and influence among individuals), further distinguishes this research from previous studies.

Chapter 3

Network Closure And Academic Attainment

3.1 Introduction

The set of outcomes where social networks play an important role is varied. It covers areas such as job searching, criminality and academic achievement. There are numerous examples of empirical investigations where network structures influence business, employment or crime opportunities (Calvó-Armengol et al. 2009; Calvó-Armengol and Jackson 2004; Patacchini and Zenou 2008). In the context of network structure it is argued that in networks where individuals are highly interlinked (close networks) there is a better access to information and more scope for norms and sanctions (Burt 2001). Network closure, the density of friendship ties, is likely to be associated with higher trust levels, better chances of establishing reputation and realisation of commitments (Koput 2010).

The main goal of this paper is to estimate the effects of network closure on academic achievement and to enhance knowledge about how this network feature may influence adolescents' human capital accumulation. Of interest is to examine whether previous closure-related findings in work or community settings apply to adolescents in social networks in schools. With the use of a unique data set, the Add Health, where ties between students are directly observed, this research aims to provide predictions regarding the impact of closure in friendship networks on individuals' attainment in education. The intuition is that closure may affect performance in education thanks to a number of influences including trust and reputation. Friendship networks in order to develop trust require closure, because in an open network, reputation is less likely to arise and collective sanctions that would ensure trustworthiness cannot be used

(Coleman 1988). Friends that know each other restrain out of norm behaviours and help to establish reputation and trustworthiness within the group under the threat of exclusion or lighter forms of punishment. Closure, by allowing for greater social sanctions between individuals through common friends, increases incentives for cooperation. More cooperation and trust should result in friends being more willing to help those lagging behind or in need of help with tests/ exams preparation, resulting in higher grade point average or better chances of going to college. Higher closure means more potential for sanctions, norms and reputation. This arguably should lead to less 'out of norm' behaviour (for example skipping classes or dropping out of high school) because deviance from resulting norms is more difficult to hide and more likely to be punished. However, there may be two sides to this story. Norms and reputations can be bad (for example the norm in a particular group may be to steal), it seems likely that closure may reinforce those norms as well. Arguably, in settings where bad norms are prevalent (for example networks of delinquent adolescents) closure may lead to adverse outcomes in education. This issue will be addressed in the empirical analysis section where the effects of closure on attainment will be estimated separately for networks of low and high quality friends (quality will be proxied by the average GPA of friends versus the school's GPA).

This investigation is related to a number of studies that focused on social networks and network structures. Sociology offers insight into network formation mechanisms and intuition on how differences in network closure may influence outcomes (Wellman and Berkowitz 1988; Coleman 1990). Network closure is argued to positively affect information flow, the level of trust and sanctions (Coleman 1988 and 1990). Through those mechanisms, thanks to increased monitoring capacity, system control, norms and commitments are more likely to fulfil.

In the economic literature a number of studies investigated the effects of network structure on job searching and education outcomes (Allcot et al. 2007; Calvo-Armengol and Zenou 2005). There exist also a vast theoretical body of literature on trust, network games and formation (Mobius and Szeidl 2009, for a survey see Jackson 2009 and 2010). Further intuition on the effects of closure on economic outcomes is offered by Granovetter (2005).

The key contribution of this investigation to the existing literature is the estimation of effects of closure in adolescents' friendship networks on academic outcomes in long-term. Of the studies mentioned above, Allcot et al. (2007) is the closest to this work. The study finds a negative relationship between community size and closure and a positive one between closure and pro-social behaviours. This study differs from Allcot et al. for a number of reasons including: focus on closure in friendship networks and long-term academic outcomes, addressing of the problem of school-specific influences and the selection of students across schools. Furthermore, the issue of possible confounding factors is carefully addressed by instrumental variable (IV) approach (using an instrument based on individual's timing of birthday celebration within a school year). In addition, the investigation into the role of closure in networks of different quality, and in establishing reputation, is expected to provide new evidence.

The chapter's structure is as follows. The next section offers a review of previous studies that investigated the effects of network structure on various outcomes including education, followed by a review of studies that dealt with closure and the issue of trust, norms, system control and reputation in social networks. The review of previous studies is followed by more information about the Add Health data and methodology used in this study; proceeded by results and discussion.

3.2 Literature review

The literature provides insight into how network closure can affect academic choices such as college going decisions or dropping out of high school. Coleman (1990) suggested a possible link between variations in network closure and a set of outcomes. In particular, he argued that networks with higher closure generate high trust between friends, which facilitates cooperation and improves welfare. The intuition is that networks with high closure allow for greater social sanctions between individuals through common friends, increasing incentives for cooperation. In line with the above logic are also arguments, given by Granovetter (2005), about channels through which closure may influence economic outcomes. The author argues that closure in networks can affect the quality and the flow of information, reward/punishment mechanisms and trust. The intuition is that in denser networks deviance from resulting norms is more

difficult to hide and more likely to be detected and punished. Glaeser et al. (2000) provide further evidence linking network closure to trust. They interpret the results as evidence that in repeated games dense social networks and trust go together. However, the arguments related to the idea of information flow also suggest that high closure, particularly in small, isolated networks, may negatively affect outcomes due to less information being acquired from outside of the friendship network and less contact with other networks (Burt 2001). This is associated with the concept of echo in information flow, mainly that that in some settings closure creates echo but not accuracy; this may have a negative effect on social capital accumulation.

Further intuition from previous research, Mobius and Szeild (2009), is that networks of high closure will lead to better outcomes, given that help from friends or helping others who lag behind in school can be considered as 'high-value' favour, a favour that requires substantial level of effort. The authors provide a plausible explanation as to why closure may be linked to schooling outcomes. The logic is that networks with greater trust (closure) allow for more social sanctions between individuals, increasing incentives for cooperation. The idea can be further illustrated by the following example: An individual is asked to offer a favour (help a friend with a school task/homework); he or she can help or refuse. Refusing in low closure setting (e.g. friends do not know each other) carries a threat of losing this particular friend that has asked for the favour. In contrast, an individual in network of high closure (most of friends know each other), in addition to the threat of losing one friend, which may be not enough to induce the individual to perform a favour, has stronger incentives to cooperate and perform high value favours by the threat of exclusion. For that reason adolescents in low closure networks could experience less backing from their peers and find schooling experience less rewarding and because of that complete less years of schooling or decide not to go to college. However, in line with previous arguments about negative norms, it is also likely that closure in specific settings (where bad norms are common) will have a negative impact on academic outcomes.

Empirical research on network closure is largely limited due to the fact that information on network closure is difficult to obtain. The data on network links is more available in more controlled settings such as project/assignment groups. Management literature offers some insight in this area. Molm (1994) argues that mutual inter-dependence

promotes cooperation, which improves group performance. Increased visibility and accountability associated with high closure counteract social loafing and therefore increase team performance (Wagner 1995). A meta-analysis of 37 studies of teams by Balkundi and Harrison (2005) suggests that teams with densely configured interpersonal ties attain their goals better and are more committed. However, some studies indicate mixed evidence or find density of a team's network of informal social ties not to affect team performance significantly (Reagans & Zuckerman 2001, Sparrowe et al. 2001).

In the economic literature the lack of suitable data is dealt with by the use of community or grade cohort size data that can arguably be treated as a proxy for closure (closure being negatively correlated with community or grade cohort size). Knack and Keefer (1997) demonstrate that an increase in country-level trust predicts a significant raise in economic growth. Numerous studies identify a negative link between school or grade size and academic outcomes (Hoxby 2000; Angrist and Lavy 1999). There is inconclusive evidence linking class size to performance in primary schools. Allcott et al. (2007) show that there exists a significant relation between school size and network closure. The study also provides some evidence that closure matters for short-term outcomes such as feeling safe, having trouble with others or most recent GPA. It is of interest to investigate the issue further, particularly, whether there are any significant academically-related long-term effects that can be attributed to network closure (does closure influence individuals' GPA in the final year of high school, years of education completed or decisions about going to college). Given that the effects of various social networks' characteristics are likely to have short and long lasting effects (Calvó-Armengol et al. 2005; Galeotti and Mueller 2005) it seems plausible that network closure will play an important role in determining adolescents' academic attainment.

3.3 Descriptive statistics and the data

The study uses data from the Add Health, which is a nationally representative sample of youths in secondary education. The fact that friendship ties are observed directly makes the data set quite unique. It is based on 144 randomly selected schools in the United States in three waves 1994 - 1995 - 2002. The design of this sample and an extensive set of questions related to everyday behaviour of adolescents, permit investigating various aspects of peer effect, social networks and their impact on academic attainment.

The network data was gathered in every school at Wave 1 (1994). Each student was asked to name up ten school friends. Because of too low response rate in 15 schools, the data set provides information on 129 of the 144 total schools (6344 students). This data is used to construct network context measures such as the density of friendship networks. The measure of network closure is provided in the first wave of the Add Health data set. It is the density of the network composed of ego and the set of ego's friends. The core sample used in this investigation excludes very small networks (up to four members, 478 cases) because the density measure is less meaningful for smaller networks and in the case of no friends is undefined. Higher than expected frequencies for density equal to .5 and 1 are due to individuals that have very small friendship networks (one or two friends). It is a common practice to exclude very small networks from analysis (this study also excludes them). The following formula was used to calculate ego-network closure:

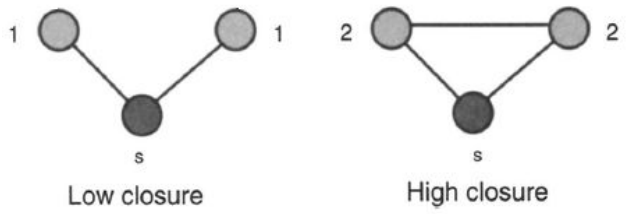
$$\text{Network closure} = \frac{\sum \mathbf{S}}{s(s-1)}$$

Where:
S = total ego send and receive network
s = number of nodes (individuals) in **S**

The total send and receive network refers to all the friendship nominations a given individual has sent and received (only direct links, friends of friends do not count).

Below two types of networks are presented, left with low closure (density) and high closure (right). Individual S in both cases has two friends; in the first one the two friends do not know each other.

Figure 3.1: Low and high closure networks.



The description of variables used in this chapter can be found in the Appendix. Table 3.1 provides descriptive statistics for selected variables that are used in the analysis that follows. The variables include: the key measure - network closure and other proxies for trust and information flow. Almost two-thirds of the respondents have college aspirations. This is noticeably higher than the actual college attendance reported in Table 2.1. The level of interaction between friends measured by the frequency of meetings is quite high. It can be seen that the individuals also spend time together outside classroom; approximately half of the sample interacts with friends during weekends.

Table 3.1: *Descriptive statistics.*

variable	mean	min	max	SD
Network closure	0.29	0.11	0.80	0.80
Network size	8.21	1.00	33.00	33.00
Commit. college	0.72	0.00	1.00	1.00
Weekend wt.friends	0.48	0.00	1.00	1.00
Time wt. friends	1.96	0.00	3.00	3.00
Information flow	3.68	1.00	5.00	5.00
Trust proxy	4.24	1.00	5.00	5.00

Note: Number of observations: 4882.

To gain a further intuition as to what may be expected from variation in closure, a comparison of networks of high and low levels is provided in Table 3.2, which reports means for selected variables by low/high closure (the split is made for density= .3 which divides the population into approximately two equal parts). The average size of a network of low closure is equal to 11 members, compared to 7 members for dense networks. Individuals in low closure networks appear to have on average lower college attendance, GPA, ability test scores and less years of education. Males seem to be more likely to have low closure networks. Accordingly to expectations based on previous studies, proxies for the levels of trust and information flow appear to be higher in close networks. Higher levels of time spent with friends in more connected networks indicate that there is more interaction with friends in those networks.

Table 3.2: *Comparison by low and high closure.*

Network closure	low	high
Variable	mean	mean
Network size	10.57**	6.55**
Network GPA	2.82**	2.86**
College	0.54*	0.61*
Years of education	7.94*	8.20*
GPA final	2.68*	2.89*
Peabody test	52.59*	56.26*
Honesty	0.53**	0.56**
Commit college	0.70	0.74
Weekend with friend	0.47*	0.49*
Timewtfriends	1.91	2.00
Information flow	3.60*	3.76*
Trust proxy	4.16	4.33
Male	0.50*	0.37*
Age	14.85**	15.07**
Grade 9	0.21**	0.15**
Grade 10	0.18	0.17
Grade 11	0.17	0.19
Grade 12	0.13**	0.18**
Steal	0.53	0.58
Non-white	0.39**	0.28**
White	0.61**	0.72**
Black	0.34**	0.18**
Hispanic	0.09**	0.10**
Asian	0.02**	0.06**
Amnative	0.01	0.02
Other	0.05	0.05
Mother edu low	0.13	0.13
Mother edu med	0.31	0.32
HH income	1119.02	1227.47
Parents' involvement	8.06*	8.49*
Urban area	0.52**	0.48**

Note: *denote statistical significance (low vs. high) at the 10% level and ** at 5% level. The split is made for closure = .3 which divides the population into approximately two equal parts.

Differences can also be identified for different races. Table 3.3 shows average closure levels for races/ethnic groups. A noticeable difference can be observed between white and black students.

Table 3.3: *Closure by race/ethnic group.*

Closure	mean
White	0.54
Black	0.36
Hispanic	0.51
Asian	0.67
American Indian	0.81
Non-white	0.41

Means reported in Table 3.2 indicate that there may be a positive link between network density and academic outcomes. One possible explanation is that staying at school as well as going to college can be linked to some form of a previously made commitment. The hypothesis based on theoretical expectations and previous studies is that networks of high density are more likely to influence commitments' realisation thanks to more scope for building up reputation (Coleman 1990). For example, in a network where everyone knows everyone else the consequences of breaking a promise are much more severe because all the friends can take actions (including the threat of exclusion). In a network of low closure, information is less likely to travel to all of an individual's friends. In an extreme case where all friends don't know each other, breaking a promise, in the worst case, can lead to the loss of one friendship. Next, a comparison is presented that may provide some intuition whether the hypothesis may be true. In Wave 1 of the Add Health data individuals were asked whether they want to go to college. It is plausible to assume that they also discussed the topic with friends or in class setting, where formal or informal commitments were made. Consequently, individuals would build up a reputation as a 'college goer' or 'no goer'. The idea is that close networks will be better for fulfilling commitments and establishing and maintaining a reputation. In this particular case, individuals are divided into two groups; those that are and are not committed to going to college (not wanting to go is assumed as no commitment because it does not require additional effort). It is expected that, on average, more commitments

will fulfil in close networks. Table 3.4 provides results that are in line with the reasoning presented above.

Table 3.4*: Commitments and actual college going rates by high and low closure.*

	Realisation of college going rates	
Closure:	High	Low
Intend to go to college	0.70	0.62
Do not intend to go college	0.32	0.34

Note: The commitment measure is reported in wave 1. College attendance is reported six years later in wave 3. The split (low / high closure) is made for density= .3 which divides the population into approximately two equal parts.

The table indicates that for individuals that do not want to go to college the average actual college attendance rate is slightly higher for those in low closure networks. Particularly interesting is the fact that there is a noticeable difference in the average college going rate for students in low and high closure networks that have made a commitment before. On average, 70% of those that have made the commitment and were in close networks went to college, compared with 62% of those in open networks. This suggests that the arguments about the relationship between network structure and commitments and reputation may also apply to adolescents in schools. At first instance, it may seem unlikely that commitments will have a significant impact six years later. However, taking into account reputation and the fact that friendships in secondary education are quite stable over time (Degirmencioglu et al. 1998) makes a lasting impact more likely. This will be further examined in the next section.

3.4 Empirical analysis

The next step is to examine the effects of closure on a set of academic achievement outcome measures including most recent GPA, years of education and college attendance. Each of these variables, except GPA, is reported in Wave 3 (year 2002) of the study. Looking at these three measures allows drawing more general conclusions and to some extent limits the possibility of omitted variables biasing the results. More formally the reduced form model is:

(1) Edu. outcomes = $\alpha + \beta \cdot$ Network closure + $\gamma \cdot$ controls + school dummies + ε

Note: Edu. outcomes include: years of education, college attendance and recent GPA. The set of controls (discussed in more detail below) includes variables listed Table 3.2 except the outcome measures and proxies for network closure (network trust, information flow).

In this equation β measures the effect of network closure on academic achievement. To obtain unbiased estimates a number of factors needs to be taken into account. Frequently, social network analysis faces various challenges that include the problem of self-selection, also discussed in Chapter 2 (for detailed discussion see Manski 1993 and 2001; Bramoulle et al. 2009), and the difficulties in capturing the true network effects. Closure appears to be a measure that is unlikely to be influenced by a large number of individual specific factors. For example, if best friends of a particular individual do not wish to form a friendship, the individual has very limited powers to influence friendship formation among them. However, the possibility should not be ignored. One option is that particular students may self-select into specific networks, for example small close networks consisting of high achievers. If this was the case, high closure networks would have tendency to be composed of college orientated individuals that study hard and have few friends that form strongly interlinked networks; as a result the coefficient of interest would be biased. To eliminate this possibility, best friends' average GPA will

be included in the analysis. Furthermore, it is important to control for an individual's ability, as it is likely to be influencing the key measures of interest. The first variable that aims to control for ability is GPA based on core subjects. The problem of unobserved ability may still persist even with GPA included. To account for that an additional ability measure, the Peabody test score (IQ proxy), will be also included in the extended set of controls.

In addition, closure is strongly linked with network size, particularly true for small networks. The ability to maintain a particular number of friends may be reflecting a specific social skill that also impacts closure and academic attainment. A possible solution to the problem is to include controls for network size. Furthermore, network composition, particularly homogeneity, may be affecting closure and outcomes of interest. In order to account for this possibility, heterogeneity measures with respect to race and age of an individual's friendship networks will be included as controls. Next, is the issue of other factors that should not be omitted from the regression as they may be linked to both, closure and outcome measures. The intuition is that higher closure may arise between individuals who have low costs of forming friendships (Hojman and Szeidl 2008; and Galeotti, Goyal and Kamphorst 2006). These low costs can arise due to individuals being close geographically, or having similar interests or other characteristics. For example, previously discussed results (Table 3.2) indicate that it is probable that preferences on education choices and social network formation vary by race/ethnic groups. Therefore it is important to account for factors such as an individual's race, parents' education, household income and location.

Furthermore, controls for individuals' physical and mental attractiveness (assessed by the interviewer) will be included. This is done because of the possibility that attractiveness could have an impact on GPA (French et al. 2009) and quite possibly on closure. Finally, school specific polices may influence the level of closure between students. For example students in particular schools may experience more interaction and higher closure levels but also better chances of going to college or staying longer in school. For that reason and to eliminate the possibility of self-selection of students across schools biasing the results, school dummies will be included in the regression. It is expected that the extensive rich set of control variables will alleviate endogeneity

problems associated with closure and that the above steps will allow estimating the unbiased effect of closure on academic achievement.

If closure is found to be affecting academic outcomes, then it is of particular interest to check whether reputation/commitments have significantly different effects in low versus high closure networks. The data allow examining whether establishing a reputation can be identified as one of the possible channels. More precisely, it is of interest to investigate if in close networks reputation is more likely to be established and commitments to fulfil. Particularly, the aim is to assess whether there is a significant difference in the effect of commitment on outcomes between low and high closure groups. The following equation will be estimated:

(2) Prob (Attend college) = $\delta + \theta \cdot$ Commitment $+ \gamma \cdot$ controls$+ \upsilon$

Note: College attendance is observed six years after a commitment is made. The commitment is going to college in the future, reported in wave 1. This will be estimated using logistic regression. Controls as in equation 1.

The hypothesis is that θ will be significantly smaller in low closure networks, due to less scope for establishing reputation, sanctions, norms and cooperation leading to lower levels of commitment fulfilments. When closure is present, in a friendship network, friends may use sanctions to limit deviations from the expected behaviour or reputation.

The analysis begins with regressing equation (1) for the three outcome variables. Table 3.5 presents estimates for the long-term outcomes listed previously, plus a short-term outcome (most recent GPA score) in column 1 and 2 (GPA). In all regressions controls for the number of friends are included. The key difference between column 1 and the remaining columns is that it does not include controls for schools' specific effects.

Table 3.5**:** *Effect of closure on outcomes.*

Outcome:	1.GPA		2.GPA		3.YOE		4.College		5.College	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
Net. Closure	0.22	*0.04*	0.19	*0.09*	0.011	*0.01*	1.01	*0.00*	1.02	*0.04*
Male	-1.07	*0.00*	-1.10	*0.00*	-0.27	*0.00*	0.62	*0.00*	0.61	*0.00*
Non-white	-0.32	*0.00*	-0.20	*0.14*	0.09	*0.24*	1.12	*0.40*	1.49	*0.01*
Mother edu low	-0.94	*0.00*	-0.92	*0.00*	-1.06	*0.00*	0.19	*0.00*	0.24	*0.00*
Mother edu med	-0.49	*0.00*	-0.40	*0.00*	-0.55	*0.00*	0.45	*0.00*	0.51	*0.00*
HH income	0.00	*0.47*	0.00	*0.97*	0.00	*0.66*	1.00	*0.91*	1.00	*0.64*
Urban	-0.10	*0.28*	0.13	*0.43*	-0.06	*0.54*	0.87	*0.43*	1.00	*1.00*
Peabody test									1.01	*0.00*
Member any									1.14	*0.00*
Steal									0.90	*0.03*
Parents' involv.									1.15	*0.00*
GPA recent					0.57	0.00			3.60	*0.00*
GPA Network					0.23	0.00			1.44	*0.00*
Col.Commitment									1.99	0.00
Add. controls	no		no		no		no		yes	
School dummies	no		yes		yes		yes		yes	
Net. size dum.	yes		yes		yes		yes		yes	
R-squared	0.0999		0.1585		0.457		0.218		0.325	

Note: Number of observations: 4822. All the columns include students' grade indicators. Column 4 and 5 report the exponent of the coefficients. Additional controls also include: Physical and personality attractiveness, motivation, health and controls for networks' delinquency and diversity with respect to race and age.

Columns 1 shows results very similar (in terms of significance and the size of the coefficient) to Allcot et al. (2007), however after accounting for schools' characteristics the impact of closure on most recent GPA becomes insignificant at the 95% level of confidence (column 2). This result supports previous arguments that school specific influences and students' selection across schools do matter and should not be omitted in this type of analysis. This finding also suggests that in short-term the effect of closure on GPA is only weakly significant. Columns 3 to 5 assess the long-term effects of closure on academic attainment. They include controls for both, the number of friends and school specific effects. Results in column 3 indicate that after controlling for the characteristics of individuals, networks, parents and location; having best friends that know each other results in more years of education. Similarly, column 4 reports exponents of coefficients for logistic regression where going to college is the outcome

of interest. In line with theoretical expectations, for a given individual, the probability of going to college six years later increases with closure of his/her friendship network.

Next, due to the worries that omission of factors such as unobserved ability could potentially lead to biased estimates, a number of variables is added to the analysis. Column 5 contains additional controls (often omitted in previous studies) that are likely to be associated with both closure and the college going outcome (including: individuals' IQ and criminal behaviour proxy, participation in extracurricular activities and parental involvement). The significance of the coefficient on network closure drops slightly but remains significant at the 95% confidence level. The same step is taken for the years of education measure (not reported). Similarly, the effects of closure remain significant and positive.

As part of robustness check, the sample is spilt into two parts (white and non-white individuals). This is done because of concern that race may be biasing the results beyond control due to differences in culture or attitudes that could influence formation of social networks and preferences on choices such as college going or the length of education. The results remain robust in both cases, suggesting that the effect of closure is similar for the two groups. The estimates do not differ significantly from what is presented in Table 3.5 and therefore are not reported here.

Next, the analysis focuses on testing the reputation and commitment hypothesis. Table 3.6 presents estimates of equation (2). The two columns report the effect of college commitment on actual college going (six year later). The columns show the effect for low (col.1) and high (col.2) closure networks. The effect of commitment on college attendance was larger for individuals in high closure networks. In low closure networks the effect is still significant but of a smaller magnitude. This finding supports previously discussed arguments that commitments are more likely to fulfil in high closure settings thanks to building up a reputation, more scope for sanctions and cooperation.

Table 3.6: *Effect of commitments on college attendance; by low and high closure.*

Outcome: College attendance				
Net. Closure:	Low		High	
	Coef.	P>|t|	Coef.	P>|t|
College commitment	1.88	*0.00*	2.91	*0.00*
Add. controls	yes		yes	
School dummies	yes		yes	
Network size dummies	yes		yes	
Pseudo R-squared	0.3411		0.3799	

Note: Columns 1 and 2 report exponents of coefficients obtained for logistic regression. All the columns include student grade indicators and control variables as in Table 3.5, col. 5. Number of observations: 2128 for low closure and 2116 for high closure.

Another question that arises is whether closure is universally good. This relates to the previously discussed issue of negative norms. In particular, is the effect of closure the same for good and bad friendship networks; does the quality of friends matter? The proposed measure of friends' quality is GPA. Let's consider two individuals, one in a friendship network where grades are on average higher than school's GPA and another with friends that on average do significantly worse than the school's average. Will the effect of closure be the same in both cases? Arguably, in low quality networks closure may be negatively affecting outcomes due to a number of influences. The intuition is that closure is better for norms or reputation, if those norms or reputation are bad ones (more likely in low quality networks), then closeness could be associated with worse outcomes in low quality networks. It is possible that networks consisting of low performers could punish or exclude individuals that try to do better. In other words, it may be a norm to do well in a high quality network, but it also may be a norm to do poorly in a low quality network. Equation 1 is used to test this idea. The sample is split into two parts; one where schools average GPA is greater than an individual's network GPA and one where the opposite is true. The results (Table 3.7) indicate that in the case of years of education the effects of closure are quite similar in terms of size and significance for low and high quality networks.

Table 3.7: *Effect of closure on outcomes for low and high quality networks.*

Outcome:	Final year GPA				Years of education			
GPA:	*Network ≥ school*		*Network < school*		*Network ≥school*		*Network<school*	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
Net. Closure	0.51	*0.02*	0.11	*0.61*	0.011	*0.02*	0.013	*0.02*
Network GPA	0.34	*0.00*	0.27	*0.00*	0.47	*0.00*	0.35	*0.01*
Network size dummies	yes		yes		yes		yes	
Add. Controls	yes		yes		yes		yes	
School dummies	yes		yes		yes		yes	
R-squared	0.49		0.43		0.47		0.44	

Note: All the columns include student grade indicators and control variables as in Table 3.5, col.5. Number of observations: 2128 for final year GPA and 2116 for years of education.

In the case of GPA in the final year, the positive effect of closure is only present for high quality networks. One of the possible explanations for the difference between GPA and years of education could be that norms related to staying at school are more universal and do not vary by the quality of networks. Interestingly, in both cases the effect of friends' GPA appears to be larger in high quality networks.

3.4.1 Using grade cohort variation

Because, even with the rich set of controls, endogeneity issues may still be a problem. Selection bias (individuals selecting into particular networks) is still possible. To address this, the following specification is considered:

$$\text{Edu. outcomes} = \alpha + \beta \cdot \text{Grade closure} + \gamma \cdot \text{controls} + \text{school dummies} + \varepsilon$$

Here, closure is no longer a measure of the individual's friendship network as in equation 1. It is now closure of an individual's grade cohort, the connectedness of peers that constitute a given cohort. This is an alternative measure for friendship networks closure. This measure is likely to be most meaningful in small schools, where closure of friendship networks is likely to be reflected in closure of a grade cohort. For this reason, a sub-sample of schools smaller than 500 students is selected for the analysis. This specification allows examining the effects of closure at a more aggregate level where self-selection is less likely to be a problem. The key assumption in identifying the true

effects is that variation in grade closure, in a given school, is exogenous to individual outcomes (given controls). Because of the possibility that other confounding factors could bias the estimates, the specification again accounts for unobserved school characteristics. Controls, in addition, include grade specific measures such as average grade ability, proportion of non-white students and gender ratio. The reason for this addition is the need to control for grade related influences that can be linked to both, an individual's performance and grade cohort closure. As results in Table 3.2 indicate, gender, race and ability are likely to be such factors.

The intuition again is that closure at grade cohort level will lead to more pro-social behaviours, consequently improving individuals' academic attainment. The results (Table 3.8) indicate that probability of going to college increases with closure of an individual's grade cohort. Being part of a close grade cohort also increases years of education an individual completes. This finding is in line with evidence from previous studies that link community closure to better outcomes (Balkundi and Harrison 2005).

Table 3.8**:** *Grade cohort variation and IV approach. Closure and attainment.*

Outcome:	YOE		College		College	
	Coef.	P>|t|	Coef	P>|z|	Coef	P>|z|
Closure (grade)	0.11	*0.00*	1.03	*0.05*		
Male ratio (grade)	-0.06	*0.09*	0.41	*0.48*		
Race composition (grade)	-1.36	*0.00*	0.18	*0.40*		
GPA (grade)	0.06	*0.00*	1.04	*0.00*		
Net. Closure (instrumented)					1.10	*0.02*
Add. Controls	yes		yes		yes	
School dummies	yes		yes		yes	
Network size dummies	no		no		yes	
1st stage inst. relevance						0.00
R-squared	0.46		0.30		0.36	

Note: In columns 1 and 2 the estimates are based on a sub-sample consisting of small schools only (those with less than 500 students). Column 2 reports exponents of coefficients obtained for logistic regression. All the columns include control variables as in Table 3.5 except for grade indicators. Number of observations: 2028 for columns 1 and 2 and 4822 for column 3. In column 3 closure of an individual's friendship network is instrumented by the individual's month of birthday indicator. Exponent of the coefficient reported in the column comes from a probit model. It was converted to logistic coefficients in order to make the comparison with previous logistic estimates easier; the rule of thumb of 1.6 is used for conversion (Stern 1989).

3.4.2 Instrumental variables approach.

A major disadvantage of the above approach is that closure of grade cohort and friendship networks may be different things even when only small schools are considered. An alternative approach is return to the friendship network closure measure used in equation 1 and to use instrumental variable method to estimate the effects of closure on academic attainment.

The following paragraph provides motivation for the instrument. At the beginning of every school year there is a lot of scope for new friendship formation because individuals are exposed to new classes and settings. Additionally at the beginning of every school year students have more time as tests, exams and accumulation of homework are less likely. The intuition is that students that have birthdays early in the school year will be more likely to have birthday parties and also to experience higher attendance at those parties because of that greater free time availability. Such parties are likely to increase the number of friendships between an individual's friends (using October as the reference month provides the most robust results). Given evidence from previous research (Robertson 2011, Bound and Jeager 2000) it can be concluded that after controlling for mother and household socio-economic characteristics (already included in the extended set of controls in Table 3.5), having a birthday early in the school year is exogenous to whether an individual goes to college. First stage estimates reveal that the instrument is not weak (p-value < 0.00 Table 3.8, column 3 and Table 3.9 in the Appendix). As expected, celebrating birthdays early in the school year is positively affecting the density of ties. In addition, it appears that individuals' ability and attractiveness play a significant role in determining closure.

With only one variable appearing to meet the requirements of a good instrument it is not possible to test empirically whether the exogeneity requirement is met. Results of the second stage regression, where having birthdays early in the school year, is used as instrument for closure of individuals' friendship network are reported in Table 3.8. The coefficient of interest increases, both in terms of size and significance, relative to the previous estimates (Table 3.5). This result reconfirms that having friends that know each other has a significant and positive effect on an individual's academic attainment (more precisely, a 1% increase in closure results 0.09 higher likelihood of attending

college). It can also be observed that the results of OLS estimation were subject to a downward bias. The results indicate that the estimates of main interest reported in Table 3.5 cannot be trusted due to possible selection problem and confounding factors. Nonetheless, the sing of the coefficient on closure remains positive, both, in the case of OLS an IV.

3.5 Discussion

Generally, the results show that closure of friendship networks leads to better outcomes in education; in line with predictions by Coleman (1990) that closure should result in pro-social behaviours. The findings indicate that when best friends of an individual are more interlinked, the individual is more likely to complete more years of education or go to college. This is in line with the theoretical expectations that link closure with reputation, more scope for sanctions, norms and cooperation within a network, and better outcomes. Commitments made six years earlier probably would not matter much for college going decision, however they do allow for reputation to arise, which is more likely to be long lasting and to have a long-term impact. Once an individual builds up a status as a college goer it seems plausible that it will follow him/her during high school. If best friends detect a behaviour that is not in line with general expectations, steps are taken to enforce the expected behaviour under the threat of sanctions or exclusion; in line with arguments offered by Koput (2010).

The examination of the effect of commitments offers further support for the results reported in Tables 3.5 and 3.8 and their interpretation. The analysis of the impact of closure in low and high ability networks provides only limited evidence in line with Burt (2001). It appears that in schools the impact of closure on college attendance is universal in both, low and high, ability networks. The only outcome where closure is likely to have a varying impact (with respect to quality) is an individuals' GPA. Arguably, the reason why closure is found to have no significant effect on grades when friends are of low ability is that the effects of connectedness work in the opposite directions (possibly because of a mixture of bad/good norms and reputations). It is worth emphasising that the results refer to adolescents in school settings only. The theory and previous empirical research suggests that in different settings or for other

outcomes (for example for low value favours) the effects of closure may be different (Mobius and Szeidl 2009; Mobius et al. 2010).

Some limitations in this study arise because the public use version of the Add Health does not include information about friends' ID numbers. For this reason, friendship nominations cannot be observed and alternative network measures cannot be constructed. This limits the analysis to the already provided network variables. A connected issue is related to the SAR nature of the data. Namely, the models presented in this study do not fully account for it. This requires caution when interpreting the results because if there is too much dependence in the data the resulting estimates may be biased and inconsistent. A possible solution would be to use, with the help of the unrestricted Add Health data, a SAR model, which could bring additional source of information and could be valuable for identification or improving estimation efficiency as in Lin (2009). In SAR models the previously mentioned "reflection problem" can be eliminated by relying on the non-linearity introduced by the variations in the peer measurements.

Furthermore, it is important to emphasize caution when interpreting the results obtained with the use of the IV method with one instrument. With only one variable appearing to meet the requirements of good instruments it was not possible to test empirically whether the exogeneity requirement is met. Ideally two or more instruments would be needed. If it is the case that having birthdays early in the school year affects students performance in school, the estimates presented in Table 3.8 in column 3 would be unreliable. However, it seems plausible that, given the previous section's motivation for the birthday indicator, the requirements of a good instrument are satisfied.

An area that this chapter leaves unexplored, due to the data limitations, is how robust are the results to other measures of closure. Ideally, an alternative measure of closure should also shed new light on how the structure of friendship ties affects adolescents' outcomes. It would be interesting to see whether substituting a broader measure, one that also includes ties between friends of friends, changes estimates significantly. A closely connected subject that requires further attentions is the idea of disconnected islands, networks that have high density of ties but few or no connections to other networks. Further study in this area, with particular focus on information flow, would

greatly improve general understanding about functioning of adolescents' social networks.

3.6 Conclusions

This paper provides evidence that closure in friendship networks has a significant impact on academic attainment. The investigation used the Add Health data set to examine the association between closure, commitment realisation and outcomes in education. Individuals in close networks are found to be more likely to complete more years of education and go to college (true for friendship networks and grade cohorts). The results remain robust to controlling for the number of friends and selection across schools. The issue of unobservable factors that could affect closure and academic attainment is further addressed by the use of instrumental variables approach. Using the timing of birthday celebration as instrument for closure reconfirms previous results. The findings are in line with earlier studies that found a significant relation between closure and outcomes. The key contribution of this study is that it assesses the long-term effects of closure on academic attainment. Furthermore, the analysis of commitments, closure and outcomes provides suggestive evidence that commitments are more likely to fulfil in high closure networks thanks to building up a reputation. The effects of closure on years of schooling are found to persist for both low and high quality networks. Due to the small number of relevant outcomes that the data set contains, it was not possible to test whether high closure could lead to worse outcomes in different settings, particularly where only 'low value' favours are performed, in line with the arguments provided by Mobius and Szeidl (2009) and Burt (1995).

Chapter 4

Community Composition and Participation of Adolescents

4.1 Introduction and background

Communities and social networks within them play an important role in determining many outcomes, including areas such as job searching, criminality and civic participation. Previous research shows that in racially heterogeneous communities, adults are less likely to take part in group activities (Alesina and La Ferrara 2000; La Ferrara 2002). This research aims to assess whether the same is true for adolescents in schools. There are numerous reasons why adolescents' participation matters. Taking part in voluntary group activities through increased transmission of knowledge, social norms and development of skills, may positively affect human capital, adolescents' schooling and labour market success (Putnam 1998).

Due to a number of factors this examination may provide results not in line with previous studies focused on adults' participation. One is that adults and adolescents may have different preferences regarding group membership and composition. For example, adolescents may be more likely to adjust to a multi-racial community and be willing to participate irrespectively of its racial fragmentation. Another factor is the difference in the size and type of communities. Schools or grade cohorts, when compared to neighbourhoods, are arguably more organised and offer significantly more interaction and opportunities to participate in various group activities. The data used in this research, the Add Health survey, allows looking at participation in extracurricular school activities that is most likely to have a positive effect on an adolescent's future.

The activities include participation in various organisations and clubs such as: debate team, computer club, newspaper, student council, honour society, language and math club. What further differentiates this research is that the data set makes it possible to focus the analysis on relatively small communities, limiting the possibility that there is a significant amount of sorting within a given community, for example a racially heterogeneous community having relatively homogenous schools with respect to race. In this study the issue of selection into school is carefully addressed by using within school variation between grade cohorts. The rich set of variables that the data set contains makes it possible to control for a number of factors such as specific health problems and network effects that have not been accounted for in previous studies.

It is expected that this analysis will widen general understanding of the effects of communities' racial composition on group membership. For economists, participation is an interesting subject for a number of reasons, including the role it plays in the formation of human and social capital, that have been shown to have a wide range of economic effects ranging from education, delinquency and criminal behaviour to job searching success (Jackson 2006; Calvó-Armengol, Patacchini and Zenou 2009; Ioannides and Soetevent 2007). Participation of adults is a well-investigated subject largely thanks to its importance for voting and political movements. Being part of a group is argued to also have other economic effects including transmission of knowledge and the development of trust (Benabou 1996; Romer 1986 and La Porta et al, 1997).

The role of ethnic conflict and racial composition in determining participation-related concepts was also investigated by a number of studies including Alesina, Baqir and Easterly (2000); Goldin and Katz (1999); Glaeser et al. (2000) and La Ferrara (2002). In line with the previously mentioned research by Alesina and La Ferrara (2000), is the last study by La Ferrara, looking at income inequality and the incentives to join economic groups. The investigation also finds evidence that heterogeneity at community level in Tanzania leads to lower levels of participation.

In summary, previous research suggests that community composition with respect to race and ethnicity should play an important role in determining adolescents' participation. However, it provides little insight as to whether similar arguments apply

to small and organised communities such as schools and adolescents within them. Majority of previous empirical investigations studied how the characteristics of adults impact on decision whether or not to participate in various groups.

The evidence on how adolescents' participation affects performance in school and academic attainment is mixed. This is largely due to limited data availability and endogeneity issues that will be discussed in the next section. Participation in extracurricular activities improves teamwork skills and self-discipline (Spreitzer 1994). Taking part in high school sport activities also improves chances of securing funding towards higher education tuition fees through sport scholarships offered by colleges for best performers. However, participation is likely to reduce the amount of time individuals spend on learning and studying. This ambiguity is reflected in the literature. A number of studies identifies a significant positive link between participation in school extracurricular activities, including sports, and schooling outcomes (Lipscomb 2006, Barron et al. 2000). Other studies provide evidence that after addressing the issue of unobservables the effects of participation are significantly reduced and in some cases insignificant (Edie and Ronan 2001). Rees and Sabia (2010) use instrumental variables approach (IV) to address endogeneity issues and conclude that the effects of sport participation on school performance are insignificant. However, the authors find some evidence of positive effects of sport participation on aspirations to attend college in the future.

This paper's main focus is on the characteristics of communities of adolescents and interactions within them. The following hypothesis, based on evidence from previous studies, will be tested: because individuals have a preference to interact with similar types, in a setting where the likelihood of interaction with other races is high, the levels of participation are expected to be considerably lower. In particular for memberships in groups that require a significant level of interaction.

The remaining part of this analysis is structured as follows: in the next section the data set is described and descriptive statistics are presented; followed by empirical analysis where the impact of communities' racial composition on individuals' participation is estimated; followed by summary.

4.2. Descriptive statistics and methods

The design of the Add Health and an extensive set of questions related to everyday behaviour of adolescents, permit for investigation of the effects of communities' composition on participation outcomes. This unique data set allows investigating the effects of communities' characteristics at various levels, including: neighbourhoods, schools, grade cohorts and individual friendship networks. Furthermore, the third wave (2002) contains information on whether the individuals went to college and on years of education completed. This can be used to examine the effects of various characteristics on long-term outcomes.

Table 4.1 contains descriptive statistics that aim to introduce selected variables that are used in this chapter. The variable, indicating participation in schools' extracurricular activities, takes values: 0 for those individuals that do not participate and 1 for those that reported that they are members in one or more groups (observed in the first wave of the study). The rate of membership, 79% of the individuals reported to be a member, is higher than the rates in studies where only adults are the subjects of interest. For example Alesina and La Ferrara (2000) report participation rates of 72% amongst adults in the United States. La Ferrara (2002) reports that the average number of groups per one individual was equal to 1.6. In the Add Health an individual is on average a member in 2.3 groups. These differences can partially be explained by greater availability of time that adolescents have as well as the increased variety of groups in secondary education. The questionnaire offers 33 choices of different groups (all the groups and a detailed description of other variables used in this study are listed in the Appendix). Out of these choices a number of alternative participation variables is created. The broadest category is a dummy variable equal to one if the individual is a member of at least one group/organisation/team in the school. The remaining participation variables indicate membership in more specific groups. They include membership in sport, civic and education-related groups. The issue of previously mentioned sorting problem also arises when individuals' general participation in extracurricular activities is considered; mainly that a significant level of participation may be attributed to memberships in

culture, language or ethnic groups. For example, it is possible that in a racially fragmented community individuals of a given race or ethnicity form a highly homogenous group, for instance a Mexican society. To address the issue, a participation variable that excludes membership in such groups is created *(excl. language and other clubs).* The data set also contains information about parental participation in teams, organisations and clubs; here the rates are significantly lower than in studies mentioned previously. The most likely explanation is the design of the question, which offers a much narrower choice of answers than in other studies or in the case of adolescents. Furthermore, all adults that completed this questionnaire had children in school age. This is likely to limit the amount of time available that parents have for participation.

Table 4.1: *Summary statistics.*

Variable	Mean	Std. Dev.	Min	Max
Race composition (grade)	0.34	0.33	0.00	1.00
Race diversity (grade)	0.24	0.23	0.00	1.00
Member of any group	0.79	0.31	0.00	1.00
Member civic group	0.20	0.40	0.00	1.00
Member civic and education	0.42	0.49	0.00	1.00
Member excluding sport	0.46	0.50	0.00	1.00
Member excl. languages and other	0.75	0.43	0.00	1.00
Member of other group	0.19	0.39	0.00	1.00
Member of sport group	0.52	0.29	0.00	1.00
Parental participation	0.50	0.49	0.00	1.00
Racial composition (school)	0.36	0.32	0.01	1.00
Racial composition (neighbourhood)	0.30	0.29	0.00	1.00
Health	2.10	0.90	1.00	5.00
Median Income	72650.76	198667.70	4999.00	999999.00*
Mother edu. low	0.15	0.35	0.00	1.00
Mother edu. med.	0.31	0.46	0.00	1.00

Note: number of observations = 4731, All the variables are reported at wave 1 of the study. *- top coded at 999999.

Table 4.1 also contains a set of variables related to racial and ethnic composition of grade cohorts, schools and neighbourhoods (places where individuals live or interact). The average community size is equal to 752 individuals for schools; and 1000 people or 452 housing units for neighbourhoods (not reported in the table). The measure of main interest is the racial composition of school (proportion of non-white students). In 71% of the school Non-whites are the minority. Racial composition at the neighbourhood level is already provided in the data set (dispersion in race composition). The measure

takes values from zero, for homogenous communities, to one (see the Appendix for more details). With respect to ethnicity in the neighbourhood, the data offers information only on the proportion of Hispanics. In 8.7% of the neighbourhoods the share of Hispanic population is greater than 25%. Race diversity (grade) captures the diversity with respect to three categories: White, Black and Other.

Table 4.2 presents averages for non-members (column 1) and members (column 2) to gain more insight about the characteristics of the two groups. The statistics indicate that males and individuals that have less educated mothers are less likely to participate in group activities. The table also shows significant differences between racial/ethnic groups. Particularly, individuals of Hispanic background appear to be less likely to be a member of an activity group. The same is true for individuals that classify their race as other than White.

Table 4.2**:** *Comparison by individual participation.*

	No participation	Participates
Variable	Mean	Mean
Race composition (school)	0.39*	0.33*
College	0.35*	0.64*
Male	0.53*	0.46*
Age	15.11	14.98
Median Income	72135.91	71239.79
Mother edu. Low	0.21*	0.11*
Mother edu. Med.	0.36*	0.29*
Parental involvement	7.84*	8.31*
Parental participation	0.40*	0.55*
Health	2.28*	2.02*
Peabody test	45.46*	55.38*
Non-White	0.37*	0.34*
Black	0.26	0.26
Hispanic	0.15*	0.08*
Asian	0.05	0.04
Am. native	0.02*	0.01*
Other	0.09*	0.05*

Note: * *denote significant difference between the means (no participation vs. participation) at the 5% level.* Number of observations = 4731.

Table 4.3 presents correlations between the variables of interest. Correlation between the measures of racial/ethnic composition at neighbourhood and school level is quite high. Individuals' membership is negatively correlated with school and both measures

of neighbourhood composition. Parents' membership is also negatively correlated with these measures.

Table 4.3: *Correlations, variables of interest.*

	Race composition (school)	Ethnic composition (neighbourhood)	Race composition (neighbourhood)	Individual Membership
Ethnic composition (neighbourhood)	0.15			
Race composition (neighbourhood)	0.30	0.29		
Individual membership	-0.04	-0.09	-0.05	
Parent membership	0.03	-0.08	-0.04	0.13

To gain further insight with regard to what characteristic may influence individuals' membership decisions the following specification is estimated where only individual determinants of participation are included:

(1) Prob (Individual civic participation)= $\alpha + \beta \cdot$ individual controls$+\varepsilon$

Note: Estimated using logit.

The dependant variable, *civic group participation* is binary and takes values 0 (no participation) and 1 (individuals is a member of at least one civic group). Table 4.4 presents some preliminary estimates. Individuals' participation appears to be increasing with ability and household's income. It also seems that parents' involvement and participation play a significant role in determining adolescents' decisions whether or not to become a member. The results indicate that males are less likely to be members. This result is not in line with findings from Alesina and La Ferrara (2000), where the opposite was true. This is likely to come from the differences in time availability between adolescent males and females and adults. It is quite possible that the rate of participation amongst females drops due to more responsibilities in adult life (working and taking care of children).

Table 4.4: *Individual determinants of participation*.

Participation: civic	Coef.	P>|z|
Black	0.228	0.00
Hispanic	0.218	0.13
Asian	0.234	0.14
Other	-0.383	0.32
Male	-0.576	0.00
Age	0.034	0.06
Peabody test	0.006	0.00
Health	-0.087	0.01
Lack of Energy	-0.162	0.00
Friends' GPA	0.328	0.00
Parental participation	0.161	0.01
Parental involvement	0.030	0.04
Mother's edu. medium	0.067	0.57
Mother's edu. high	0.115	0.33
HH income	0.001	0.01

Note: *first column contains logit coefficients. No. obs.* = 4731*,*

This investigation is particularly focused on the relationship between race and participation. It can be observed in Table 4.4 that after controlling for a set of characteristics, individuals classified as Hispanics and Other are no longer less likely to take part when compared to Whites. Interestingly, Black individuals are found to participate significantly more. This result was not clear when only partial correlations were considered and lower rates of participation amongst races/ethnic other than white groups were observed. This has a significant consequence for next steps in this analysis. Because Black individuals are the minority in almost all communities, the percentage of Blacks is positively correlated with the racial composition measure. This means that if individuals' group membership will be lower in racially fragmented communities, it is not due to the link between the share of black individuals and the communities' composition. What is more, because Black individuals are found to participate more, it is now less likely to find a significant effect of composition on membership, in line with arguments in Alesina and La Ferrara (2000).

The next step would be to include characteristics of communities into the model and to investigate how racial composition affects an individual's participation decisions, with the aim to determine what role community composition has in determining an individual's participation and to see how the results relate to previous studies based on adults. However, an analysis that looks at the effects of neighbourhoods' composition may provide biased results if there is a significant amount of sorting into schools within the communities. This provides motivation to concentrate on a more specific form of community, such as schools. The clear advantage of looking at small types of communities is that the sorting of individuals into formal homogenous groups is less likely to occur. It is also likely that individuals will have better knowledge about their community if it is a small one. This could arguably lead to a stronger influence of the community's composition on an individual's decision whether or not to participate. Controls in such specification would include the set of individual, household and school characteristics such as gender, age, IQ proxy (Peabody test) and household's income.

The above step would have clear limitations, because students' self-selection into schools and school specific policies may play a significant role in determining a wide set of outcomes including participation (Hoxby 2000, Bifuclo 2008, Bramoulle et al. 2009). For this reason an alternative and preferred specification that addresses the above problems by exploring the variation in racial composition within schools, is presented. This approach uses within-school variation between grade cohorts to identify effects of racial composition on participation. More formally the reduced form model is:

(2) Prob (Individual participation)= $\alpha + \beta \cdot$ grade cohort composition + $\gamma \cdot$ ind. controls + school dummies $+\varepsilon$

Note: Estimated using logit.

Here, β will measure the effect of grade cohort composition with respect to race in a given school on an individual's participation. This specification will include school dummies to control for unobservable school characteristics. In addition, controls as in Table 4.4 and grade cohort variables such as average income and ability will be included. The main identification strategy is based on the assumption that variations in grade cohort composition, within school, are exogenous to individual student outcomes, after controls are included.

4.3 Empirical analysis

Results in Table 4.5 report the effects of communities' racial composition on individuals' participation in clubs and organisations (eq.2). In the specification communities' average income, ability and gender composition are controlled for. This is the preferred specification as it addresses problems of sorting and unobserved school specific influences (for example, it limits the possibility that school specific policies and resources could play a major role in determining the outcome of interest). The key measure of interest (column 1) is the proportion of non-White individuals in a given grade cohort. In line with theoretical expectations the coefficient is negative and significant at the 95% level of confidence. Here the negative effect of racial composition on participation is noticeable (a 1% increase in the proportion of non-Whites in a community reduces the likelihood of an adolescent's participation by 0.6%). In addition, participation of parents is identified as an important determinant of adolescents' participation. Interestingly household income does not seem to play an important role. Black individuals are found to be more likely to take part in extracurricular activities.

Table 4.5: *Community racial composition and participation of individuals.*

	Col. 1:		Col. 2:		Col. 3:		Col. 4:	
Sample	*Whole*		*Whole*		*Whites*		*Non-whites*	
Outcome: Participation	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Race composition	0.381	0.042						
Race diversity			0.981	0.627	0.881	0.921	1.515	0.197
Community ability	0.982	0.061	1.007	0.403	1.013	0.252	0.661	0.992
Black	1.770	0.011	1.660	0.169			1.851	0.159
Asian	1.582	0.219	1.541	0.320			4.042	0.131
Male	1.049	0.719	0.571	0.001	0.461	0.000	0.758	0.358
Peabody test	1.006	0.020	1.011	0.000	1.011	0.005	1.011	0.048
Parental involvement	1.065	0.051	1.032	0.022	1.135	0.000	1.003	0.753
Parental participation	1.832	0.000	1.702	0.002	2.212	0.000	0.971	0.933
Mother edu. Med.	0.928	0.722	0.791	0.373	0.701	0.315	0.988	0.980
Mother edu. High.	1.640	0.035	1.283	0.384	1.262	0.519	1.196	0.721
HH income	1.003	0.510	1.002	0.681	1.001	0.772	1.006	0.510
School dummies	Yes		Yes		Yes		Yes	
Pseudo R^2	*0.09*		0.13		0.14		0.14	

Note: The columns contain exponents of logistic regression coefficients. No. obs. = 4731 (col. 1), 3286 (col.2), 1914 (col. 3) 1372 (col. 4) . In addition to variables from Table 4.4, controls include: male/female ratio, average income for and grades, school dummies, urban area indicators.

In order to capture the effects of community composition with respect to other races (not only the non-white proportion), an alternative community composition measure is introduced. Column 2 of the table contains results where the effect of variation in racial diversity of grade cohort within a given school is measured. The coefficient of main interest decreases in terms of significance. Again, in addition to individual controls, this specification also includes grade cohort controls such as ability, income and gender

ratio. The benefit of using the race diversity measure is that it is sensitive to differences in racial composition other than the white and non-white differences. This is important particularly in the case of schools where whites are in minority. The disadvantage of using this measure is that individuals' race is not always correctly reported. In a significant number of observations no specific race is given or multiple races are selected. This will be further discussed in the next section.

Next, it is of particular interest to investigate whether there are any differences in the impact of composition on participation that depend on whether an individual is white or non-white. To check this, the sample is split into two: one consisting of only white individuals (column 3) and the other of non-white students (column 4). The coefficient of primary interest is insignificant in both cases. The same can be observed in the case of the previous measure of community composition (the proportion of non-whites, not reported here). One probable explanation for this is that the sample is significantly reduced in the case of the split. For example for the non-white sample it can be observed that the coefficients on other variables also become insignificant with the exception of the IQ proxy (Peabody test).

The above findings are linked with only one outcome measure – participation in at least one club or organisation. This motivates the next step of this analysis. Mainly, it is of interest to find out whether the effects of racial composition are different for memberships in different types of groups. Previous research provides some evidence and insight on what to expect. According to a model developed by Alesina and La Ferrara (2000) the level of interaction and excludability of groups should play an important role. The intuition is that in groups where a significant level of interaction between members is required, the effect of racial composition will be stronger. Similarly, the negative effect should be more noticeable if the level of excludability of a given club is low.

Table 4.6 reports only coefficients of the key variable of interest obtained from the same regression as in Table 4.5 (column 1, eq. 2). Each row represents a separate regression with a different participation measure. The findings mostly agree with the above intuition. Civic participation for example is likely to be characterised by low levels of excludability and significant levels of interaction between members. The non-White

ratio is found to have a significant effect on civic participation. The first row excludes participation in language and other clubs because these clubs are likely to have formal or informal membership requirements, such as an ability to speak a particular language or in the case of various nationality clubs, being of a given nationality/origin. The significance of the coefficient and its size increased when compared to estimates from Table 4.5. Similarly some clubs such as honour society can only be joined if an individual's school performance (GPA) is high. The non-White ratio is also found to have a significant effect on participation that excludes performance-related groups, which also can be linked to low excludability and high interaction. On the other hand, sports clubs offer a good example of high excludability where specific skills may be required. The coefficients for sports and languages participation are not statistically significant. The last row indicates that the measure of key interest has no significant effect on the number of activities an individual is involved in.

Table 4.6: *Participation by type of group.*

Participation:	exp	p-value
All excl. languages and other	0.353	0.014
All excl. performance liked	0.433	0.054
Civic	0.382	0.025
Civic and education related	1.052	0.880
Languages	1.524	0.400
Sport	1.028	0.930
Other club	1.073	0.866
Number of activities	0.896	0.700

Note: Column one contains exponents of logistic regressions coefficients.
No. obs. = 4731. All regressions include controls as in Table 4.5, Col.2.

It is possible that the regressions underestimate the impact of composition on participation. This can be due to a small percentage of schools where the non-Whites are the majority (where a particular race other than White becomes a majority). In this case increases in the non-White ratio measure would indicate increased likelihood of interaction with similar individuals, arguably leading to increased participation. This could lead to underestimation of the racial composition effect in the above examination. In order to account for this possibility the above regressions are re-estimated with

schools, where the problem is likely to be present, being excluded from the analysis. This does not affect the results significantly.

4.4 Discussion

This investigation provides evidence suggesting that racial composition of communities affects adolescents' participation in extracurricular activities in line with Alesina and La Ferrara (2002) study of adults' participation. The most likely explanation of the results is the preference of individuals to interact with similar types (Jackson 2009b). In addition, the contribution of this analysis is that problems related to sorting within communities, and selection into schools, are carefully addressed.

Still, there remain many issues associated with this this analysis that need to be highlighted. One of the limitations of this investigation arises due to the lack of information about the intensity of individuals' participation. Many individuals report participation in unlikely high number of clubs. Future study should take into account how intense or frequent is the participation. A connected issue that may also affect the reliability of the presented estimates is associated with the fact that a significant number of clubs, even after the exclusion of language and other clubs, may still have some form of formal or informal membership requirements (for example ethnicity or race). Some of the extracurricular activities are not clearly defined in the Add Health data set. It is difficult to pinpoint what is hidden under the *other club* category. Arguably, the exclusion may not be the best solution and alternative ways should be considered. Furthermore, there still remains the possibility that some clubs are quite homogenous with respect to race. Ideally one would like to check the composition of clubs with respect to race before proceeding to further stages of analysis. Similarly, due to the nature of the survey, a precise identification of an individual's race is an issue. It can be stated that if this was not the case, a more detailed study of the effects race diversity would be possible.

In addition, it needs to be highlighted that the estimation strategy relies on the assumption that, given controls, grade cohort variation in race composition is exogenous

to an individual schooling outcomes. If this assumption does not hold the presented estimates would be unreliable. Moreover, in line with the issues presented in the previous chapters, it is important to emphasise caution when interpreting the results of this analysis because of the discussed issue of the spatial nature of the data. Ideally, the SAR nature of the data should be accounted for.

The low number of outcomes that can be related to the intensity of interaction between individuals in the public use data set is somehow constraining. One possible direction for future study is to investigate how the characteristics of communities studied here impact on other outcomes such as non-extracurricular school activities. For further research it is also of interest to measure to what extent the intensity of interaction, in a diverse setting, influences participation decisions.

4.5 Summary

The investigation used the Add Health data set to examine the association between communities' composition with respect to race and participation in extracurricular school activities in secondary education. Results of this analysis show that the composition of adolescents' communities matters for individuals' participation outcomes in extracurricular school activities. The results find further support in previous research connecting community heterogeneity to reduced propensity to participate in social activities including education (Alesina and La Ferrara 2000). The findings support the previously discussed argument that the effect of composition is stronger in groups or activities where interaction between individuals is likely to be more frequent. This is also in line with the idea that individuals have a preference to interact with similar types (Jackson 2009b). This research opens many avenues for further research. Of particular interest would be to examine the role of frequency or the intensity of interactions in determining participation decisions.

Bibliography

Akerlof G. and Kranton R. 2000: Economics and identity. *Quarterly Journal of Economics*. 715–753.

Alesina A. Baqir R. and Easterly W. 2000: Redistributive Public Employment. *Journal of Urban Economics*, Elsevier, vol. 48, 219-241.

Alesina A. and E. La Ferrara 2000: Participation in Heterogeneous Communities. *Quarterly Journal of Economics*. 847-904.

Alesina A. and E. La Ferrara 2002: Who Trusts Others? *Journal of Public Economics*, 85, 207-234.

Allan G. 1989: Friendship: Developing a sociological perspective. *Journal of Social and Personal Relationships*.

Allcott H. Karlan D. and Möbius M. 2007: Community Size and Network Closure. *The American Economic Review*, Vol. 97, No. 2 May, 2007) 80-85 Published by: American Economic Association

Angrist J. and V. Lavy 1999: Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement. *The Quarterly Journal of Economics*, MIT Press, vol. 114(2), 533-575, May.

Arcidiacono P. and S. Nicholson 2004: Peer Effects in Medical School. *Journal of Public Economics* 89:327-350.

Arcidiacono P. and Vigdor J. L. 2010: Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College. *Economic Inquiry*, Western Economic Association International, vol. 48(3) 537-557.

Arrow K. J. 1999: Observations on Social Capital. In Dasgupta and Serageldin, 3–5.

Averett S. L. Argys M., Rees D. 2011: Older siblings and adolescent risky behavior: does parenting play a role? *Journal of Population Economics* 24(3): 957-978.

Babcock P. 2008: From Ties to Gains? Evidence on Connectedness and Human Capital Acquisition. *Journal of Human Capital*, University of Chicago Press, vol. 2(4), 379-409.

Balkundi P. & Harrison D.A. 2005: Ties, leaders, and time in teams: strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*.

Barron J. M. Ewing B. T. and Waddell G. R. 2000: The effects of high school athletic participation on education and labor market outcomes. *The Review of Economics and Statistics*, 82(3), 409-421.

Benabou R. 1996: Heterogeneity, Stratification, and Growth. *The American Economic Review,* 4311, National Bureau of Economic Research.

Bifulco R. Fletcher J. M. Ross Stephen L. 2011: The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 25-53(29)

Bonacich P. 1987: Power and Centrality: A Family of Measures. *American Journal of Sociology* 92:1170-1182.

Bound J. Jaeger D. 2000: Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? in (ed.) *(Research in Labor Economics, Volume 19)*, Emerald Group Publishing Limited,83-108.

Bramoulle Y. H. Djebbari, B. Fortin 2009: Identification of peer effects through social networks, *Journal of Econometrics*, Volume 150, Issue 1, May 2009, 41-55.

Burt R. and A. V. Peterson 1998: Smoking cessation among High School Seniors. *Preventive Medicine* 27, 319–327.

Burt R. D. and K. T. Dinh 2000: Predicting Adolescent Smoking: A Prospective Study of Personality Variables. *Preventive Medicine* 30(2): 115-125.

Burt R.S. 2001: Bandwidth and echo: trust, information and gossip in social networks', in A. Cassella and J. Rausch (eds), *Networks and Markets: Contributions from Economics and Sociology*, New York: Russell Sage.

Burt R.S. 1995: *Structual Holes, The social structure of competition*. 1st ed. Cambridge, Massachusetts, and London, England: Harvard University Press

Calvo-Armengol A and Zenou Y. 2005: Job matching, social network and word-of-mouth communication. *Journal of Urban Economics*, Volume 57, Issue 3, May 2005, 500-522

Calvó-Armengol A. and M. O. Jackson 2004: The Effects of Social Networks On Unemployment and Inequality. *American Economic Review*, 94(3), 426-454.

Calvó-Armengol A. Patacchini E. and Zenou. Y. 2009: Peer Effects and Social Networks in Education and Crime. *The Review of Economic Studies*, Volume 76, Number 4, October 2009 , 1239-1267(29)Mimeo. Universitat Autonoma de Barcelona.

Canals J. Bladé J. et al. 1997: Smoking and personality predictors in young Spanish people. *Personality and Individual Differences*. 23(5): 905-908.

Chang M. J. Witt-Sandis. D. and Hakuta. K. 1999: The Dynamics of Race in Higher Education: An Examination of the Evidence. *Equity and Excellence in Education*. 32(2). (1999). 12-16

Coleman J. S. 1988: Social Capital in the Creation of Human Capital. *The American Journal of Sociology* 94:S95-S120.

Coleman J. S. 1990: Foundations of Social Theory. Cambridge, MA: Harvard University Press.

Cook P. J. and Hutchinson, R. 2007: Smoke Signals: Adolescent Smoking and School Continuation in Marina Bianchi (ed.) The Evolution of Consumption: Theories and Practices. *Advances in Austrian Economics*, Volume 10, Emerald Group Publishing Limited, 157-186

Cook T. D. Deng Y. & Morgano E. 2007: Friendship influences during early adolescence: The special role of friends' grade point average. *Journal of Research on Adolescence*, 17, 325–356.

Cook P. J. MacCoun R. Muschkin C. and Vigdor J. 2008: The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management*, 27: 104–121.

Cotterell J. 2007: Social Networks in Youth and Adolescence. *Adolescence and Society Series.*

Degirmencioglu S. M., K. A. Urberg, J. M. Tolson, P. Richard 1998: Adolescent Friendship Networks: Continuity and Change over the School Year. *Merrill-Palmer Quarterly*, Vol. 44,

Di Pasquale D. Glaeser E. 1999: Incentives and social capital: Are homeowners better citizens? *Journal of Urban Economics* 45 (2), 354-384.

Digest of Education Statistics 2001: National Center for Education Statistics. http://nces.ed.gov/pubs2002/2002130.pdf.

Ding W. L. and S. F. Lehrer 2007: Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics* 89:300-312.

Eide E. and Ronan N. 2001: Is participation in high school athletics an investment or a consumption good? Evidence from high school and beyond, *Economics of Education Review*, 20, issue 5, 431-442

Eysenck H.J. 1991: Smoking, personality, and stress: psychosocial factors in the prevention of cancer and coronary heart disease. New York: Springer-Verlag,

French M. T. and P. K. Robins, et al. 2009: Effects of physical attractiveness, personality, and grooming on academic performance in high school. *Labour Economics* 16(4): 373-382.

Galeotti A. Goyal S. and Kamphorst J. 2003: Network Formation with Heterogeneous Players, *Economics Discussion Papers* 562, University of Essex, Department of Economics.

Galeotti A. and Mueller G. 2005: Friendship Relations in the School Class and Adult Economic Attainment. *IZA Discussion Papers* 1682. Institute for the Study of Labor (IZA)

Glaeser E. L., D. I. Laibson, J. A. Scheinkman, and C. L. Soutter 2000: Measuring Trust. *Quarterly Journal of Economics*.

Glaeser E., D. Laibson, J. A. Scheinkman, and C. L. Soutter, 2000: What Is Social Capital? The Determinants of Trust and Trustworthiness. *Quarterly Journal of Economics*, 811–846.

Goldin C. and Katz L. 1999: Human Capital and Social Capital, The Rise of Secondary Schooling in America, 1910 to 1940. *Journal of Interdisciplinary History*. 29: 683-723.

Goux D. and E. Maurin. 2007: Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal* 117:1193-1215.

Granovetter M. 2005: The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives*, 19(1): 33–50.

Haller A. and George W. Ohlendorf. 1970: The Educational and Early Occupational Attainment Process: Replication and Revision. *American Sociological Review* 35:1017-27.

Hallinan M. T. 1991: School Differences in Tracking Structures and Track Assignments. *Journal of Research on Adolescence*. 1. 3. 251-275.

Hanushek E. A. and Gomes-Neto, J. & Leite, R. & Frota-Bezzera R., 1997: Health and schooling: Evidence and policy implications for developing countries, *Economics of Education Review*, Elsevier, vol. 16(3), 271-282, June.

Hanushek E. A.. John F. Kain. Jacob M. Markman. and Steven G. Rivkin. 2003: Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics* 18:527-544.

Hattie J. A. C. 2002: Classroom composition and peer effects. *International Journal of Educational Research* 37(5): 449-481.

Haynie D. L. 2002: Friendship Networks and Delinquency: The Relative Nature of Peer Delinquency. *Journal of Quantitative Criminology* 18:99-134.

Henry G. T.. and D. K. Rickman. 2007: Do peers influence children's skill development in preschool? Economics of Education Review 26:100-112.

Hersch J. and Pickton T. S. 1995: Risk-taking activities and heterogeneity of job-risk tradeoffs, *Journal of Risk and Uncertainty*, 11, 205–17.

Hojman D. and Szeidl, A., 2008: Core and Periphery in networks. *Journal of Economic Theory* Volume 139, Issue 1, March 2008, 295-309.

Hoxby C. 2000: The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics* 115:4, 1239-1285 Online publication date: 1-Nov-2000.

Ioannides Y. and Soetevent A. 2007: Social Networking and Individual Outcomes beyond the Mean Field Case. *Journal of Economic Behavior and Organization*. Volume 64, Issues 3-4, November-December 2007, 369-390

Jackson M. O. 2010: An Overview of Social Networks and Economic Applications, July 2010, in the The Handbook of Social Economics, edited by J. Benhabib, A. Bisin, and M.O. Jackson, Elsevier Press.

Jackson M. O. 2009: Networks and Economic Behavior. *The Annual Review of Economics*, Vol. 1, 489-513, January 2009.

Jackson M. O. 2009b: Social Structure, Segregation, and Economic Behavior, Available at SSRN: http://ssrn.com/abstract=1530885

Jackson M. O. 2006: The Economics of Social Networks. in Proceedings of the 9th World Congress of the Econometric Society. Vol.1. Cambridge: Cambridge University Press.

Kang C. H. 2007: Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics* 61:458-495.

Kézdi G. and Z. Cseres-Gergely 2008: Why are extraverted young men less likely to receive higher education? Evidence from Hungary. Working paper, Central European University

Knack S. and P. Keefer 1997: Does social capital have an economic payoff? A cross-country investigation, *Quarterly Journal of Economics*, vol. 112, no. 4, pp. 1251-1288.

Koput K. W. 2010: Social Capital: An Introduction to Managing Networks Edward Elgar Publishing.

Kubicka L., Matejcek Z. Dytrych Z. and Roth Z. 2001: IQ and personality traits assessed in childhood as predictors of drinking and smoking behaviour in middle-aged adults: a 24-year follow-up study. *Addiction*, 96: 1615–1628.

La Ferrara E. 2002: Inequality and Group Participation: Theory and Evidence from Rural Tanzania, *Journal of Public Economics*, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

La Porta R. Shleifer A. and Vishny W. et al, 1997: Trust in Large Organizations, American Economic Review, *American Economic Association,* vol. 87(2), 333-38, May.

Xu Lin 2010: Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables, *Journal of Labor Economics*, University of Chicago Press, vol. 28(4), pages 825-860, October.

Linville P. and Jones E. 1980: Polarized Appraisals of Outgroup Members. *Journal of Personality and Social Psychology*. 38: 689–703.

Lipscomb S. 2006: Secondary school extracurricular involvement and academic achievement: A fixed effects approach. *Economics of Education Review*, 26(4), 463-472.

Liu X. & Patacchini E. & Zenou Y., 2011: Peer Effects in Education, Sport, and Screen Activities: Local Aggregate or Local Average?, *CEPR Discussion Papers* 8477, C.E.P.R.

Mannix E. and Neale M. 2005: What Differences Make a Difference? *Psychological Science in the Public Interest.* Vol. 6. No. 2 (October 2005): 31-55.

Manski C. 1993: Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*. LX(1993). 531-542. American Economic Association.

Manski C. 2000: Economic Analysis of Social Interactions. *Journal of Economic Perspectives*. XIV. 115–136.

Mason D. A. and Burns, R. B. 1995: Teachers' views of combination classes. *The Journal of Education*, 89(1), 36–45.

Mason D., & Burns R. 1996: Simply no worse and simply no better may be simply wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research*, 66(3), 307–332.

Medved R. 1966: Body height and predisposition for certain sports. *The Journal of sports medicine and physical fitness.* Phys Fitness. 1966 Jun;6(2):89–91.

Miller L.S. and E.E. Garcia. 2004: Better Informing Efforts to Increase Latino Student Success in Higher Education. *Education and Urban Society* 36(2):189-204.

Mobius M. & Szeidl A. & Ambrus A. , 2010: Consumption Risk-sharing in Social Networks, *Economics Working Papers* 0079, National Bureau of Economic Research. NBER

Mobius M. and Szeidl A. 2009: Trust and Social Collateral, The Quarterly Journal of Economics, MIT Press, vol. 124(3) Harvard University; National Bureau of Economic Research. NBER

Molm L. D 1994: Dependence and Risk. Transforming the Structure of Social Exchange. *Social Psychology Quarterly* Vol. 57, No. 3, Special Issue: Conceptualizing Structure in Social Psychology (Sep., 1994), pp. 163-176.

Nahapiet J. & Ghoshal S. 1998: Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2), 242-266.

NCES [National Center for Education Statistics, U.S. Department of Education] 2005: Youth Indicators, 2005: Trends in the Well-Being of American Youth.

O'Brien E.M. and C. Zudak. 1998: Minority-Serving Institutions: An Overview. New Directions for Higher Education 102:5-15.

Patacchini E. and Zenou Y. 2008: The strength of weak ties in crime, *European Economic Review*, Elsevier, vol. 52(2), pages 209-236, February.

Patacchini E. and Zenou Y. 2006: Racial Identity and Education, *CEPR Discussion Papers* 5607, C.E.P.R

Persico N. Postlewaite and D. Silveman 2004: The effect of adolescent experience on labour market outcomes: the case of height, *J. Political Economy* 112 (2004) (5).

Putnam R. and S. Yonish 1998: New Evidence on Trends in American Social Capital and Civic Engagement: Are we Really Bowling Alone? *Journal of democracy*.

Reagans R. & Zuckerman E. W. 2001: Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science,* 12: 502–517.

Rees I. Argys M. Averett L. and Witoonchart, B. 2006: Birth order and risky adolescent behavior. *Economic Inquiry*, 44: 215–233.

Rees D. I. and Sabia J. 2010: Sports Participation and Academic Performance: Evidence from the National Longitudinal Study of Adolescent Health. *Economics of Education Review*, Vol. 29, No. 5, pp. 751-759.

Rees D. I. and Sabia J. 2011: The Effect of Migraine Headache on Educational Attainment. *The Journal of Human Resources,* Vol. 46, No. 2, pp. 317-332.

Richard O. and Shelor R. 2002: Linking top management team age heterogeneity to firm performance: Juxtaposing two midrange theories. *International Journal of Human Resource Management*.13. 958–974.

Riordan C. and Shore L. 1997: Demographic Diversity and Employee Attitudes: Examination of Relational Demography within Work Units. *Journal of Applied Psychology*. 82: 342–58.

Robertson E. 2011: The effects of quarter of birth on academic outcomes at the elementary school level. *Economics of Education Review* 30(2): 300-311.

Romer P. M. 1986: Increasing Returns and Long-run Growth, *Journal of Political Economy*, University of Chicago Press, vol. 94(5), 1002-37, October.

Sacerdote B. 2001: Peer Effects with Random Assignment: Results for Darmouth Rommmates. *Quarterly Journal of Economics*. 681-704.

Sewell W. H. Archibald O. Haller. and G. W. Ohlendorf 1970: The Educational and Early Occupational Attainment Process: Replication and Revision. *American Sociological Review* 35:1017-27.

Sparrowe R.T. Liden R.C. Wayne S.J. & Kraimer M.L. 2001: Social networks and the performance of individuals and groups. *Academy of Management Journal*, 44: 316-325.

Spreitzer E. 1994: Does participation in interscholastic athletics affect adult development*?* *Youth and Society*, 25, 368-387.

Steinberg L. Brown B. & Dornbusch S. 1996: Beyond the classroom: Why school reform has failed and what parents need to do. New York: Simon and Schuster.

Stern S. 1989: Rules of thumb for comparing multinomial logit and multinomial probit coefficients. *Economics Letters* 31(3): 235-238.

Tajfel H. (ed.) 1982: Social Identity and Intergroup Relations. Cambridge: Cambridge University Press.

Triandis H. Kurowski L. and Gelfand M. 1993: Workplace Diversity. In Triandis. H.. Dunnette. M. and Hough. L. (eds) *Handbook of Industrial and Organizational Psychology*. Vol. 4. Palo Alto. CA: Consulting Psychologists Press. pp. 769–827.

Veenman S. 1995: Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, 65(4), 319-381

Veenman S. 1996: Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, 66(3), 323-341.

Wagner J. A. 1995: Studies of individualism collectivism: Effects on cooperation in groups. *Academy of Management Journal,* 38: 152-172.

Wellman B. and Berkowitz, S. D. 1988: *Social Structures: A Network Approach*. New York: Cambridge University Press.

Wolaver A. 2008: Effects of heavy drinking in college on study effort, grade point average, and major choice. *Contemporary Economic Policy*. Volume 20 Issue 4, Pages 415 - 428

Zenger T. & Lawrence B. 1989: Organizational demography: The differential effects of age and tenure distributions on technical communications. *Academy of Management Journal*. 32.353–376.

Zimmerman D. J. 2003: Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics* 85:9-23.

Appendix

Table 3.9: *Reduced form model of network closure. First stage regression.*

Network closure	Coef.	p-value
Birthday indicator	0.027	0.001
Drinks alcohol	-0.005	0.394
Drugs user	0.001	0.894
Male	-0.017	0.001
Steal	-0.003	0.477
Physical Attractivenes	0.010	0.010
Member any	0.012	0.130
Peabody test	0.001	0.019
Mother edu. low	0.009	0.351
Mother edu. med	-0.009	0.112
Parental involvment	0.001	0.258
HH Income	0.001	0.046
School dummies	yes	
R squared	0.35	
First stage relevance test, f-stat.		0.001

Note: *N. of obs.:* 4822. *Additional controls as in Table 3.5*

Description of variables:

Outcome variables:

GPA final year: Grade point average refers to the average of grades in the final year of high school, based on transcripts. GPA is the mean of grades across four core subjects. Grades are weighted as follows: A = 4, B = 3, C = 2, D =1 and F = 0.

Recent GPA: Grade point average refers to the average of grades in the same year when friendship data is gathered, based on transcripts. GPA is the mean of grades in English. Grades are weighted as follows: A = 4, B = 3, C = 2, D =1 and F = 0.

College: Dummy variable equal to 1 if the respondent was or is attending college, 0 otherwise.

Years of Education: Number of years of education completed including primary, secondary and higher education.

Individual Participation variables: set of group membership variables that the individual is involved in schools:

> Member any, general participation: dummy variable =1 if the individual is a member of at least one group (organisation/team/society/club) at school. The following choices were available: French, German, club Latin, club, Spanish club, Book club Computer club, Debate team, Drama club, Future Farmers of America, History club, Math club, Science

club, Band, Cheerleading/dance team, Chorus or choir Orchestra, Other club or organization, Baseball/softball, Basketball Field hockey, Football, Ice hockey, Soccer, Swimming, Tennis, Track, Volleyball Wrestling, Other sport, Newspaper, Honour society, Student council, Yearbook .

Member civic group: dummy variable =1 if the individual is a member of at least one of the following: Debate team, Future Farmers of America, History club, Newspaper, Student council, Yearbook.
Member civic and education: dummy variable =1 if the individual is a member of at least one of the following: Debate team, Future Farmers of America, History club, Newspaper, Student council, Yearbook.
Member excluding sport: French, German, club Latin, club, Spanish club, Book club Computer club, Debate team, Drama club, Future Farmers of America, History club, Math club, Science club, Newspaper, Honour society, Student council, Yearbook . Other club or organization
Member excluding languages and other: dummy variable =1 if the individual is a member of groups as in *Member any* excluding membership in language groups and other groups that are not listed.
Member of other group: dummy variable =1 if the individual is a member of other group. Not specified.
Member of sport: dummy variable =1 if the individual is a member of any sport club/ team.
Member of languages: dummy variable =1 if the individual is a member of any language related club.
Number of activities: total number of activities the individual is involved in. Top coded at 10.

Variables of key interest:

Age diversity: assesses heterogeneity of the respondent's network with respect to age (taking values from 0 to .84), missing if ego is the only member of the underlying ego network, or if all members of the ego network (including ego) have missing data on the attribute. If all members of the ego network who have valid data on attribute A share the same trait, heterogeneity =0, This measure is based no send and receive nomination.

Age diversity (sent): the same as above except that this measure is based on sent nominations.

High age diversity: Dummy variable equals to 1 if individual's network is highly diversified (Age diversity equal or greater than .5)

Younger network: dummy variable =1 if the average age of an individual's friendship network is e lower than individual's age (one year or more).

Older network: dummy variable =1 if the average age of an individual's friendship network is significantly higher than individual's age (one year or more).
Avg. netw. age≠individual's age: dummy variable =1 if the average age of an individual's friendship network is significantly different than individual's age.

Older best friend: dummy variable =1 if the individual has an older best friend (older if the difference in the month of birth is greater than 12 months)

Younger best friend dummy variable =1 if the individual has a younger best friend (younger if the difference in the month of birth is greater than 12 months)

Best friend of different age: dummy variable =1 if the individual has a best friend of age different than his/her own (different if the difference in the month of birth is greater than 12 months)

Network Closure (friendship network density): Density of the network composed of ego, the set of alters nominated by ego, and the set of alters who nominated ego. The density of the respondent's network is equal to the ratio of the number of links (nominations) and the number of friends (it takes values from 0 to 1), the measure is missing if ego is the only member of the underlying ego network. The measure increases with connectedness of ego's friends.

Closure (grade): Average density of an individual's grade cohort. (taking values from 0 to 1)

College commitment: dummy variable equal to 1 if the individual has made a college going commitment in wave 1 of the study, 0 otherwise.

Race composition (grade): Proportion of students in an individual's grade cohort that are non-White.

Race diversity (grade): Diversity with respect to races: White/Black/Other. Constructed the following way:

$$C = \frac{k(N^2 - \sum f_i^2)}{N^2(k-1)}$$

where k is the number of categories/races, N^2 is the sum of all categories squared, and $\sum f_i^2$ is the sum of squared category frequencies over all i (=1,k) groups. If D = 0, only one category is nonzero; if D = 1,all category frequencies are equal.

Instrumental variables:

Incidence of Smoking: number of days the individual smoked cigarettes in the past 30 days.

Birthday indicator: Dummy variable equal to 1 if the individual celebrates birthday early in the school year (October). Equal to 0 for other months.

Additional variables:

Individual variables:

Age: Age of the respondent, measured in years in 1994 (wave 1).

BMI: body mass index of the individual measures at wave 1. BMI= kg/m^2.

Drinks alcohol: dummy variable equals to 1 if drank alcohol in past month.

Drugs user: dummy variable equals to 1 if used drugs (including LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills, without a doctor's prescription) in the past month.

Grade 7 -12: Dummy variables indicating current school grade of the respondent, 10th grade is the reference category.

Health: Set of health variables including:

General health assessed by the individual on 1 to 5 scale (1= excellent ,5 = poor),
Lack of energy: How strongly do you agree or disagree with each of the following statement: I have a lot of energy. (1=strongly agree, 5=strongly disagree).
Frequency health problems: In the last 12 months, how often did a health problem occurred (4=everyday, 0=never)
Health affecting activities: In the last month, how often did a health or emotional problem cause you to miss social or school activity (4=everyday, 0=never):

Injury: A 1 to 5 scale indicating how severe was the worst injury the respondent had in the past year. 1 = very minor 5 = extremely serious.

Male: dummy variable = 1 if individual is male.

Motivation: Individual's level of motivation based on individual's self assessed motivation toward schoolwork. 0=very motivated 4=unmotivated.
Never smoked: dummy variable equal to 1 if the individual never smoked cigarettes.

Non-white: dummy variable =1 if the race of respondent is non-white.

Peabody test: measures the respondent's individual's ability (IQ proxy). Adjusted and standardised for ages. The variable represents Cross-sectional Percentile Rank from Wave 1 (1994).

Personality Attractiveness: Measures how attractive in terms of personality is the respondent compared with other adolescents of {HIS/HER} age? (Assessed by the interviewer). 1=very unattractive, 5=very attractive.

Physical Attractiveness: Measures how physically attractive is the respondent compared with other adolescents of {HIS/HER} age? (assessed by the interviewer). 1=very unattractive, 5=very attractive.

Puberty: Measures how physically mature was the respondent compared with other adolescents of HIS/HER age? (assessed by the interviewer). 1=very immature, 5=very mature.

Race/group Variables: A set of dummy variables including Asian, Black, Hispanic, White Non-white, American Indian, equal to 1 if individual is of a given race/group , otherwise equal to 0.

Expected age: Dummy variable equal to 1 if the individual is of expected age for given grade level.

Seatbelt: reported seatbelt use when driving or riding in a car 0=never, 4=always.

Smoked in 30 days: dummy variable=1 if the individual smoked cigarettes in the past month.

Steal: dummy variable equals to 1 if individuals stole something in the past year.

Time with friends: Variable measuring intensity of an individual's interaction with friends in a week before interview (hanging out with friends). 0=no time, 3= five or more times.

Trust proxy: self-reported measure of how much the individual feels that his/her friends care about him; treated as a proxy for trust that the individual enjoys. 0= not at all, 5=very much.

Weekend with Friend: Dummy variable equal to 1 if the individual spent time with a best friend during the past weekend.

Family variables:

HH income: Total household income in thousands of $, before taxes in 1994 (wave 1). Includes the income of everyone in the household, and income from welfare benefits, dividends, and all other sources.

Mother's education: dummy variable(low and medium=benchmark) based on Mother education. Low refers to less than high school. Medium = high school High= above high school.

Parental Participation: dummy variable =1 if the individual's parent is a member of at least one group (organisation/team/society/club) including Military veterans organization, Labour union, Hobby or sports group, such as a bowling team or a ham radio club, Civic or social organization, such as Junior League, Rotary, or Knights of Columbus.

Parental involvement: increasing scale 1 to 10, 1= not at all 10=very much. How much do you feel that your parents care about you?

Sibling indicators: set of dummy variables:

S*iblings:* dummy variable = 1 if individual has siblings, 0 otherwise.

Older siblings: dummy variable =1 if has older siblings, 0 otherwise.

Network variables:

Friends' GPA: the average GPA of the individual's best friends (wave 1).

Information flow: this measure is a proxy for the information flow in the network. The individual is asked how many of his/her close friends knew about most recent romantic relationship when it began.

Total Network Centrality: Ego's centrality (position in the total school network), weighted by the centrality of those to whom he/she sends ties (Bonacich 1987). Increases with centrality.

Network delinquency: number of friends in an individual's network that stole something in the last year.

Network heterogeneity race: heterogeneity of the respondent's network with respect to race (taking values from 0 to 1), missing if ego is the only member of the underlying ego network, or if all members of the ego network (including ego) have missing data on the attribute. If all members of the ego network who have valid data on attribute A share the same trait, heterogeneity =0

Network size: the number of friends that are nominated ego and who nominate ego, plus ego.

Neighbourhood and community variables:

Community ability: the average Peabody Test score in school or grade cohort.
Community income: the average household income in school or grade cohort

Ethnic composition (neighbourhood): Proportion Hispanic , dummy variable = 1 if Hispanic population share in the neighbourhood where the individual lives is grater than 25%

Race composition (neighbourhood) : dispersion in race composition(White/Black/Other), of the neighbourhood where the individual lives.

GPA (Grade): Grade point average of an individual's grade cohort, refers to the average of grades in the final year of high school, based on transcripts. GPA is the mean of grades across four core subjects.

Male ratio (grade): Proportion of students in an individual's grade cohort that are males.

Median income: Median neighbourhood household income (in thousands of $). Based on U.S. Bureau of the Census defined geographic area, which in 1990, averaged 452 housing units, It is the lowest level of geography for which the Census Bureau publishes sample data, and thus captures the most localized available contextual characteristics of the areas in which individuals live.

Race composition (school): Proportion of students in an individual's school that are non-White.

Urban area: dummy variable equals to 1 if the respondent lives in urban area.