From Ties to Gains?
Evidence on Connectedness and Human Capital Acquisition

ABSTRACT

This paper uses micro-level data on social networks in middle and secondary schools to estimate effects of connectedness on education attainment outcomes. The analysis addresses concerns about unobserved neighborhood and school-level heterogeneity by using within-school variation between grade cohorts to identify effects of connectedness. Main findings include that being part of a more connected cohort within a given secondary or middle school is associated with significantly higher years of schooling attained and higher probability of having attended college, 7 years later. (JEL codes: J24, J15).

Philip Babcock
University of California, Santa Barbara

July, 2008

ACKNOWLEDGEMENTS

This work benefited from comments and conversations with Kelly Bedard, Hoyt Bleakley, Julian Betts, Vince Crawford, Julie Cullen, Roger Gordon, Gordon Hanson, Peter Kuhn, Mindy Marks, Valerie Ramey, James Rauch, and Joel Sobel, and from comments by Caroline Hoxby, Susan Dynarski, Steven Rivkin, Jeffrey Smith, and others attending the NBER Education Program Meeting, Fall, 2006. 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 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).

I. INTRODUCTION

Economists have long been aware that social networks may have a significant role to play in explaining economic outcomes. One challenge for empirical researchers has been to find micro-level measures of connectedness or social interaction that are sufficiently subtle to test nontrivial hypotheses. A second and perhaps more serious challenge has been to find sources of exogenous variation in social networks so that inferences may be drawn. This paper attempts to address these challenges using network data from the National Longitudinal Survey of Adolescent Health (Add Health). The paper estimates the effect of connectedness (as measured by friendship ties in middle and secondary school) on education attainment and the relationship between racial heterogeneity and connectedness.

Does the connectedness of a school, or of a grade within a school, influence human capital acquisition? The peer effects literature provides a point of departure. While most recent studies find that an influx of high performing peers improves an individual's performance on standardized tests (Betts, Zau, and Rice [2003]; Hoxby [2000]; Hanushek, Rivkin, Markman, and Kain [2002]), the specific mechanism has not been clearly identified. The standard technique in recent studies has been to estimate the mean response by individuals to small exogenous changes (e.g., from students migrating in and out of specific schools, or from natural variations in gender composition by cohort) in the mean performance of their peers. There are indications, however, that mean peer performance does not tell the whole story. Hoxby [2000] finds mean peer performance to be an inadequate summary statistic for influence of racial composition on individual outcomes. Angrist and Lang [2002] find that busing minority students into

Boston public schools did not significantly alter the performance of white peers in the receiving schools but that the performance of minority third graders did appear to be affected. All peers, it would seem, are not equal. A number of current policies derive from the assumption that adding high-performing peers to low-performing schools or classrooms will benefit low-performers, unambiguously. This may not be the case. By using characteristics of the peer network to explain long-run outcomes—rather than mean peer characteristics—this paper attempts to address and extend the peer effects literature.

A second strand of literature relating to connectedness is the social capital literature.[1] A number of previous studies have used self-reported measures of "trust" as a proxy for connectedness or have relied on cross-city or cross-country variation in measures of trust or connectedness for identification.[2] In general, this literature suggests that measures of trust are associated with positive economic and social outcomes. However, data on friendship networks allow for a more micro-based approach to connectedness than has been possible in cross-region studies. New questions may be posed, stricter quantitative methods applied, and the skill acquisition decision problem addressed from an uncommon perspective.

A main finding of this analysis is that connectedness, as measured by links between one's schoolmates, is associated robustly with long-run education outcomes. Specifically, being part of a more connected grade cohort within a given secondary or middle school is associated with significantly higher years of schooling attained and higher probability of having attended college, 7 years later. Also, racial heterogeneity of

[1] For clarity, exposition in this paper avoids the term "social capital." The term has different meanings in different settings. "Connectedness," as defined in Section III, captures with greater specificity the relevant notion here.
[2] See Alesina and La Ferrara (2004).

school or grade appears negatively associated with connectedness. The remainder of the paper is organized as follows: Section II motivates the analysis; Section III describes the data and the empirical strategy; Section IV analyzes empirically the relationship between connectedness and long run skill acquisition; Section V explores racial heterogeneity and school size as determinants of connectedness; Section VI summarizes and draws conclusions.

II. CONCEPTUAL FRAMEWORK

A. Overview

Concerns about connectedness inform some recent education policies. Policies promoting the construction of smaller schools, for example, rely in part on the conviction that small schools allow students to "know their classmates and teachers better than they could in a larger school."[3] It is argued that this causes students to connect more strongly to the school and to acquire more human capital. But connectedness is rarely measured in any direct or rigorous way to back up this assertion.

A relevant theoretical framework is presented in Akerlof and Kranton [2003]. They posit a model in which skill acquisition outcomes for utility-maximizing students depend, among other things, on the degree to which they connect to the goals or ideals of the school. The model allows characteristics of the school or its student population to influence the effort cost of attending school and thereby alter skill acquisition choices. An empirical challenge motivated by the model is to find institutional characteristics related to effort cost that can be well measured. Connectedness to an ideal, for example, would

[3] Linda Shaw, The Seattle Times, November 5, 2006. See also Bill and Melinda Gates Foundation, Annual Report, 2003, www.gatesfoundation.org.

seem difficult to measure. However, it is possible to measure students' connections to each other. One could imagine that the effort cost of acquiring skill might be lower in a more-connected, less factionalized environment and that utility-maximizing students might choose to stay in school longer in these settings, other things equal. This could be the case because students help each other learn more effectively in a connected setting. It could also be because students find connected environments to be more pleasant (increased "social utility"). The analysis here, then, attempts to examine a specific institutional characteristic—the connectedness of the student population—and investigate how changes in this characteristic may relate to skill acquisition choices.

B. A Model of Skill Acquisition

A simple model helps formalize the relevant intuitions. The analysis focuses on education as consumption good. Consider a model in which agents value both academic rewards and the social rewards of school attendance, and individuals vary in their academic and social abilities. *A* is the utility due to academic accomplishment. *S* captures utility associated with social interactions. Effort may be directed toward production of academic accomplishment, e_a, or toward the production of social rewards, e_s. Agents are heterogeneous in their ability endowments, θ_a, θ_s, and production increases with effort input ($f_{e_a} > 0, g_{e_s} > 0$). Here, θ_a and θ_s raise the agent's production for the (internally produced) commodities *A* and *S*, respectively ($f_{\theta_a} > 0, g_{\theta_s} > 0$). Agents maximize:

$$
\begin{aligned}
&\text{Max}_{e_a,e_s} U(A,S) \\
&\text{s.t. } e_a + e_s = 1 \\
&e_a \geq 0, e_s \geq 0 \\
&A = f(e_a, \theta_a) \\
&S = g(e_s, \theta_s)
\end{aligned} \qquad (1)
$$

If there is an interior solution, the utility gained from a marginal unit of effort devoted to producing the academic consumption good must equal the marginal benefit associated with effort directed toward producing the social consumption good:

(2) $$U_A f_{e_a} = U_S g_{e_s}\ .$$

$\underline{U}$ is the utility derived from the outside option—dropping out of school and taking a low skill job. (In general, $\underline{U}$ may depend on θ_a, θ_s, as abilities and tastes for activities in school may relate to abilities and tastes in the workplace.)

Figures IA and IB show the feasible consumption set and indifference curves for high-θ_s agents and high-θ_a agents respectively. The figure depicts a situation in which academic utility and social utility are perfect substitutes. (Linear utility in the figure is for simplicity of exposition.) One could imagine a change in the composition of schoolmates that altered the opportunity cost of making friends or of academic rewards. The usual assumption in the peer effects literature is that high-endowment classmates make it easier for teachers to teach effectively and for low-endowment students to learn. An influx of new students with high parental education endowments (high θ_a, as a result) would raise f_{e_a} at every level of effort for their peers. (Marginal gains in academic accomplishment require less effort.) Evidence in Section V will indicate that students may acquire friends more easily when the school or grade-cohort population is more homogeneous. An influx of new students that reduces homogeneity might then lower g_{e_s}, as more effort may be required to make or to keep friends, or to navigate a factionalized environment. It is not hard to imagine both these changes occurring simultaneously, as depicted in Figure IC and ID. The marginal rate of transformation of academic utility for social utility falls. As drawn in IC, the decrease in g_{e_s}/f_{e_a} lowers total utility for the low-θ_a individual to an amount below the level of the outside option, $\underline{U}$. Low-θ_a types (who would have stayed in school initially) drop out. As drawn in Figure ID, high-θ_a types who initially would have dropped out experience an increase in utility and stay in school.[4]

It is possible to draw the production sets so that these outcomes do not obtain. Stronger assumptions on functional forms are needed to generate the effect unambiguously. If $c_1\theta_a$ and $c_2\theta_s$ are the constant marginal products f' and g', respectively, then the production set of internally produced commodities is non-convex and agents will specialize (in the case that utility is linear). Here c_1 and c_2 are meant to be suggestive of institutional factors that influence the ease with which academic and social utility are produced. In Figure IE and IF, $f = e_a c_1 \theta_a + d_1$ and $g = e_s c_2 \theta_s + d_2$. Because the starting point is an all-social utility corner solution in Figure IE, a marginal increase in c_1 (which increases the marginal product of effort for academic utility), will not cause substitution into academic effort and will not increase utility, whereas a marginal decrease in c_2 (which decreases the marginal product of effort for social utility) will reduce utility, unambiguously. High-θ_s, low-θ_a types, in this example, are made worse off when c_1 rises and c_2 falls, while high-θ_a, low-θ_s types (Figure IF) are made better off. The assumptions used in this example would seem rather strong; however, linear utility and constant marginal products were not strictly necessary. More generally, non-convexity of the

[4] One could argue that having higher social ability raises rewards both at school and in the workplace (i.e., that $\underline{U}$ varies positively with θ_s). An agent with high social ability would not necessarily then be more likely to stay in school. However, any exogenous change that made the individual more productive at acquiring friends in school while leaving unchanged the individual's capacity to acquire friends out of school would increase the likelihood that the individual stayed in school. It is institutional changes of this type that motivate the model and the empirical investigation.

production possibility set may lead to specialization by some agents, if there is strong substitutability between A and S.[5]

C. Stylized Predictions

One measure of S will be the number of friends the individual nominates or is nominated by—proxied in some settings, by the number of friendship links in a school or grade cohort. If production of S rises, holding A constant, utility stays the same or rises. If utility rises then the probability that the individual stays in school also rises.[6] Several stylized predictions, suggested by the modeling exercise, can be taken to the data:

1) Agents with higher S will stay in school longer (controlling for A).

2) If connectedness in an institution increases g_{e_s} (i.e., reduces the effort required to produce social utility), then agents in connected institutions should stay in school longer.

3) If there is specialization, production of S will generally determine the utility of low-θ_a types. Factors that hinder the production of A will not generally alter utility or influence the decision to stay in school. (This prediction is more fragile, as described above.)

III. DATA AND EMPIRICAL STRATEGY

[5] It is also possible that A and S could be complements. But because time must ultimately be divided between social and academic activities, and because of the large body of sociological evidence on identity and social groups in school settings, the account here will emphasize results that obtain if A and S are substitutes. Sociological evidence would seem to imply that agents face a choice between defining themselves as academic utility producers and social utility producers, as peers impose punishments for effort choices that deviate from group norms associated with these types. See Akerlof and Kranton [2002] for a summary of this evidence.

[6] This assumes that observed S is not positively correlated with the value of the outside option. If a high observed S indicates that an indivdual forms friendships easily, then this could indicate the individual would derive high utility from low-skilled work, as well (social interactions being a component the work setting). I assume here that the marginal benefit of increased social ability in school (where the agent would typically be exposed to hundreds of potential friends, long periods of time devoted to socializing, dances, proms, athletic events, numerous clubs and organizations) is larger than in a low-skill job.

A. Data

The Add Health survey, conducted by the Carolina Population Center from 1994 to 2002, consists of data on adolescents in 132 schools across the country, grades 7-12. The in-school portion of the Wave 1 survey, conducted in 1994-1995, contains cross-section data on about 90,000 adolescents. Also, school administrators filled out questionnaires describing characteristics and policies of the schools in the sample. A subset of the initial sample, about 20,000 subjects, was selected for the in-home portion of Wave I. This second and more extensive set of interviews with students and parents took place in 1994-1995. The Carolina Population Center interviewed the in-home subjects again in 1995-1996 (Wave II), and again in 2001-2002 (Wave III). When appropriate weights and cluster coefficients are used, regressions on data from each of the surveys, or from merged samples, yield results representative of the U.S. population. Wave III measures of skill acquisition include years of schooling, college attendance and labor force participation about 7 years after the original Wave I surveys. These will be the dependent variables in the regressions that follow.

The most interesting aspect of the Add Health Survey, for the purpose at hand, is the data on friendship networks. Respondents in the in-school survey nominated up to 10 friends from the school roster, 5 male and 5 female.[7] The analysis here will be limited to the 113 schools in which at least 50% of the total student population filled out in-school questionnaires. For most of these schools, 75% or more of the entire student population filled out questionnaires. Thus, it is possible to see the general size and structure of friendship networks in these institutions. Figure II.A depicts as a directed graph the

[7] Some nominees were from a sister school. All school connectedness measures were constructed from same school nominations only.

friendship network for one such school. UCInet and Netdraw were used to create this diagram.[8] Netdraw adjusts the shape of the diagram so that the physical distance between the nodes, as displayed, rises with the number of links on the shortest path between the nodes. Nodes far apart on the graph tend to have higher degrees of separation. Clustering of nodes in the diagram, then, indicates clustering in the network. In Figure II.A, the differing shapes of nodes represent different races. The startling feature of this graph is the extreme clustering of connections along racial lines. Very few links connect members of different racial enclaves. Though this is an extreme case, it offers a first suggestion that mean responses to mean peer performance could be misleading.

B. Connectedness

Sociologists have constructed numerous measures of network centrality and connectedness. I focus here on several of the simplest: An agent's connectedness is the number of friends she has. The in-degree of a student is the number of friendship nominations she received from other students in the school. In Figure II, a student's in-degree is visible as the number of arrows pointed toward her node. A student's out-degree is the number of students she nominated as friends. This is the number of arrows from her node to other nodes. The main measure of an individual's connectedness used in the analysis that follows is the number of friendship links with respect to which the individual is either the sender or receiver, the sum of in-degree and out-degree.[9]

[8] See Borgatti, Everett, and Freeman[2002].
[9] If the agent derives utility primarily from friends with whom she interacts on a regular basis, then these would seem the appropriate measures. The emphasis here is on social utility, rather than on the flow of information across the structure of a network. Other measures of centrality and connectedness capture network characteristics relevant to the latter by counting friends of friends across multiple degrees of separation and/or weighting well-connected friends more heavily than less-connected friends. (See

In Section IV, "links" or connections will be the regressors of interest. A first strategy is to regress long–run education and employment outcomes directly on individual connectedness, as measured by links to and from other agents.

3) $$Y_{ij} = \beta_1^1 C_i + \beta_2^1 X_{ij} + \beta_3^1 S_j + \beta_4^1 R_j + \varepsilon_{ij} .$$

Here, i indexes individuals, j indexes schools, C_i is individual connections, X_{ij} is a vector of individual characteristics, S_j is a vector of school characteristics, R_j is a vector of region, state, or city characteristics, and ε_{ij} is the individual-specific error term. The potential for unobserved individual heterogeneity motivates a second specification:

4) $$Y_{ij} = \beta_1^2 C_{-i(j)} + \beta_2^2 X_{ij} + \beta_3^2 S_j + \beta_4^2 R_j + \varepsilon_{ij} .$$

The regressor of interest, $C_{-i(j)}$ (the construction of which will be described in Section IV), is the connectedness of the school, absent the individual's ties. There remains a concern that unobserved differences between schools or neighborhoods could be confounded with connectedness in the analysis. To address this concern, a final (preferred) specification controls for unobserved school and neighborhood heterogeneity by including school dummy variables, D_j, characteristics of grade-level peers, G_{jg}, and by defining connectedness at the grade level,

5) $$Y_{ijg} = \beta_1^3 C_{-i(g)} + \beta_2^3 X_{ijg} + \beta_5^3 G_{jg} + \sum_j \alpha_j D_j + \varepsilon_{ijg} ,$$

The main identification strategy, then, relies on the assumption that variations in grade cohort connectedness, within school, are exogenous to individual student outcomes, given the controls. Possible sources of endogeneity (such as common environment, within grade) will be discussed in detail in Section IV.

Wasserman and Faust[2000] for descriptions.) The direction of the sign on connectedness in the following regressions is robust across alternative measures of connectedness, though precision of the estimates varies.

Table I shows descriptive statistics of measures that will be used as right-hand variables in regressions of long-run skill on friendship ties. The average number of links to and from an individual in the sample is 8.74. The lower portion of Table I contains measures from the Add Health dataset that will be used as control variables. Covariates at the individual level include age, grade, sex, parental education, race, grade point average, a "new student" [10] dummy variable, and (for the smaller In-home sample) parental income, score on the Add Health vocabulary test, and average distance from the student's residence to the residences of schoolmates.[11] Recent work by Alesina and Laferrera [2003] and others motivates a focus on measures of heterogeneity in discussions of connectedness, in particular racial heterogeneity at the school level. The heterogeneity variable used here is a Herfindahl index subtracted from 1:

$$H_j = 1 - \sum_k {r_{jk}}^2 \tag{6}$$

where r_{jk} is the fraction of the population of race k in school j and the racial categories are White, Black, Asian, American Indian, and Hispanic. Covariates at the school level also include school size, average class size, proportion male, PTA participation rates, percent new teachers, a breakdown of the school's population by race, average parental education, standard deviation of parental education, average GPA, and the fraction of new students. Covariates at the county level include racial heterogeneity (a Herfindahl index, as for schools), median income, standard deviation of income, and crime rate. State minimum wage may be a measure of the attractiveness of dropping out of school and has been included as well.

[10] New students are students in their first year at the school.
[11] Students from remote neighborhoods may be less apt to make friends in school. See the discussion in Section IV.

IV. CONNECTEDNESS AND LONG-RUN OUTCOMES

A. School-Level Results

Is connectedness correlated with long-run skill acquisition? Table II, a first pass, displays findings from regressions of long-run skill-related outcomes on individual connectedness ("links") and covariates. I use total number of friendship ties as the measure of connectedness in the preferred regression; however, results for In-degree and Out-degree are very similar, and will be reported and discussed in Section IV.B. Results from OLS and Probit regressions of years of schooling, college participation and labor force participation[12] on connectedness and controls appear in columns 1, 2 and 3 of Table II. Estimates of the coefficient on links are positive and significant. Other coefficients have the expected signs; higher family income leads to higher skill-related outcomes, as do higher parental education and higher scores on the Add Health vocabulary test.

If a person who reports having numerous friends defines friendship in a less restrictive way than does a person who reports few friends, the willingness to nominate friends could be associated with advantageous personal traits such as optimism and confidence. These traits may drive the observed correlation with measures of long-run skill. To test for this, I ran the regressions using In-degree—the number of friendship nominations received—as the regressor of interest. The results were qualitatively very similar—a strong indication that subjective definitions of friendship do not drive the correlation.[13]

[12] Some respondents were still in school at the time of the Wave III surveys. Labor force participation regressions include only those respondents who were not still in school at the time of the Wave III surveys.
[13] These results, reported in Table IV, will be revisited in the discussion of robustness in Section IV.B.

It could also be that innately gregarious or charismatic individuals experience long-run outcomes indicative of higher skills because social ability is itself a skill.[14] In Columns 4, 5, and 6, school-level connectedness replaces individual connectedness as the regressor of interest. Individuals attending a friendlier institution might derive more immediate utility and stay in school longer, regardless of their individual endowments of gregariousness or charisma.

Consider a school with *n* students. One measure of school connectedness for individual *i* in school *j* is

$$C_{-i(j)} = \frac{\sum_{k \neq i, l \neq i} X_{kl}}{10(n-1)} \tag{7}$$

where *X* is an *n* X *n* matrix, X_{kl}=1 if *k* nominated *l* as a friend and 0 otherwise. The numerator is the number of links in the network minus the number of links to or from node *i*. Because students may nominate a maximum of 10 friends, the maximum number of links in the network is 10*n*. If we imagine removing one student from the network, the maximum number of links would be 10(*n*-1).[15] $C_{-i(j)}$, then, measures the connectedness of the network, absent *i*'s links. Multiplying this measure by 10 yields a measure, $\widetilde{C}_{-i(j)}$, with a more straightforward interpretation: the average number of links per student (absent *i*'s links) for *i*'s schoolmates. Columns 4, 5, and 6 in Table II use this second measure.

Clearly, school-level unobservables could be driving the results in Table II. Table III addresses this concern by including school-level dummy variables for all

[14] If (in the workplace) social skill were a complement to skills acquired through education, then one could imagine individuals with higher initial endowments of social skill choosing more schooling because the return to schooling would be higher for these individuals.

[15] The maximum of 10 allowed friendship nominations was not so low as to make the data uninformative. Only about 3% of the respondents nominated 10 friends.

specifications. Further, in Columns 4, 5, and 6 of Table III, grade-cohort connectedness is the regressor of interest (and grade-level control variables replace school-level controls.) Though the effect of connectedness on labor participation in Column 3 goes away in Column 6, the positive and statistically significant correlation between connectedness and education outcomes persists across specifications. This yields a major result of the paper: *Having been part of more connected grade cohort within a given school is associated with higher levels of schooling attained and greater probabilities of attending college in the long run.* School and neighborhood specific unobservables, then, do not appear to drive the correlation between connectedness and observed education outcomes.

The magnitudes of the connectedness coefficients would seem economically significant as well as statistically significant. Table III, Columns 4, 5—the preferred specifications—indicate that an increase of 1 link in the average links per student in *i*'s grade cohort is associated (about 7 years later) with an increase of .12 years of schooling acquired by *i* and an increase of 1.7 percentage points in the probability that *i* attended college. (Here, the latter is the weighted population mean of the estimated marginal effects in the "college" probit regression.) Within-school standard deviation of grade cohort connectedness is .87. Thus, the association in Column 5, if causal, implies that a 1-standard-deviation increase in grade-cohort connectedness raises average years of schooling by a tenth of a year and the probability of college attendance by 1.5 percentage points. These effects would seem large enough to be of interest to parents, policy-makers, and median voters.

B. Robustness and Alternative Explanations

1. Instrumental Variables

An alternative strategy is to use grade-cohort connectedness as an instrument for individual connectedness. Indeed, the coefficient on individual connectedness could then be interpreted as the marginal increase in education outcomes associated with having 1 more friendship tie. Estimated coefficients for the IV regressions are reported in Columns 7, 8, and 9. For the instruments to be valid, however, grade-cohort connectedness would have to influence the individual solely through its effect on individual friendship ties. It is not clear why this should be true. The individual could derive greater utility from being part of a more connected cohort, even if the individual did not acquire more friends herself. An awareness of tensions between one's schoolmates, for example, could create anxiety, reducing social utility without altering the individuals' friendships. If this were the case, the instrument would be correlated in a direct way with the outcome variables. Because the exclusion restriction is problematic here, the IV specification is not emphasized. The main findings are in columns 4 and 5: Being part of a more connected cohort appears to influence skill acquisition outcomes. The emphasis is on regressions in Table III that leave open the channel through which this influence works.

2. Alternative Measures of Connectedness

Table IV, which contains results of 36 separate school-fixed-effects regressions, shows that the main findings above are robust to alternative specifications. Columns 1 and 4 display estimates of the coefficient on connectedness in school fixed-effects regressions for several different measures of connectedness. Rows 2 and 3 show that nominating more friends and being nominated by more friends are both associated with significantly higher education attainment. More important, lines 5 and 6 indicate that being part of a more connected grade cohort within a given school—where connectedness may be defined by either of these measures—is associated with higher education outcomes. The positive association between connectedness and long-run education outcomes persists for all 6 connectedness measures in Table IV.

Clubs and extracurricular activities provide an additional measure related to connectedness. The vitality of clubs and extracurricular activities may be a school characteristic that decreases the effort cost of producing social utility, which by the logic of the model should be associated with increased education outcomes. Regressions analogous to those in Table II, Columns 4 and 5, available upon request, do show that a higher average number of clubs participated in by one's schoolmates is associated with an individual acquiring significantly more years of schooling.

3. Simultaneity

Columns 2, 3, 5, and 6 of Table IV contain results for additional regressions that test robustness of the main findings. Agent *i*'s friend-making behavior may influence others in her school or grade cohort, causing her friends to form more friendships with third parties. If this were the case then *i*'s own friendliness would explain, in part, the connectedness of *i*'s school or cohort (excluding *i*'s links). This is akin to the "reflection problem," as formalized in Manski[1993]. To mitigate this problem, I excluded small schools from the sample. If there are only a few students in *i*'s school then *i*'s charisma could plausibly influence the average number of friendships *i*'s schoolmates form with third parties. But if there are numerous students in *i*'s school, then it would seem less plausible that *i*'s friendliness influences significantly the number of ties between *i*'s schoolmates and third parties. As reported in Table IV, results for all 6 measures of connectedness were robust to excluding schools with fewer than 100 students or excluding schools with fewer 250 students (though precision of the estimates in the college attendance regressions falls with decreased sample size.) Main results do not appear to be driven by small schools.

4 Peer Effects and Cognitive Ability

Grade-specific unobservables are a major concern in the preferred specification. Connectedness of one's grade cohort could simply be a proxy for having a peer group with high cognitive ability. Friendship data in the Add Health survey are based on students filling out a questionnaire in which they write in names of their friends and find codes for those names on a list of ID numbers for students in their school. One might worry that students who are better at taking tests would also be better at following instructions, more likely to fill out the friendship questionnaire more completely—and thus would nominate more friends. In such a case, friendships would be an indirect measure of cognitive skill, and correlated with education attainment. Do students stay in school longer in connected grade cohorts because their peers are of higher ability? Or do they get more schooling because increased social utility makes staying school more attractive than the outside option? It is worth distinguishing these channels.

Columns 4, 5 of Table III—the main results in this section—regress individual education outcomes on grade level connectedness. The regressions include peer characteristics as covariates, but one could go further. Students in the Wave I sample completed the Add health vocabulary test. If the connectedness of an individual's grade cohort is primarily a proxy for cognitive skill, then one would expect a positive correlation between test scores on this vocabulary exam and the connectedness of the grade cohort. Similarly one would expect a positive correlation between the individual's GPA and the connectedness of the grade cohort. School fixed effects regressions of Add Health vocabulary test score on grade cohort connectedness (with or without the full set of controls) yield negative and statistically insignificant estimates of the grade-cohort connectedness coefficient, as do similar regressions of individual GPA on grade cohort connectedness. Both are summarized below[16]:

$$\text{AH Vocab} = \underset{(.233)}{-.047}\ \text{Links (grd)} + \text{school dummies}$$

$$\text{GPA} = \underset{(.020)}{-.023}\ \text{Links (grd)} + \text{school dummies}$$

Grade cohort connectedness is not associated with higher individual test scores or with higher GPAs—a finding which would seem to cast doubt on the notion that connectedness of grade cohort is merely a proxy for cognitive skill.

It is counterintuitive, perhaps, that students from more connected grade cohorts appear to stay in school longer but do not appear to score higher on measures of cognitive skill. Previous research suggests that this is plausible, even if counterintuitive. Some institutional changes may cause students to connect to the education setting in a way that makes them willing to stay in school longer, *even absent gains in performance*. Findings from the Tennessee STAR experiment (Krueger and Whitemore, 1999) show test scores

[16] These include only school dummies as controls. Results for regressions that include the full set of covariates are similar.

of students exposed to small classes in kindergarten through third grade exceeded those of students in large classes. However, test score gains disappeared by the end of high school. In spite of this, eight years after they had left the experiment, minority students who had been in the small classes were significantly more likely to take college entrance exams than those in control groups. Changes in attitude toward education, rather than changes in performance, would appear to be the enduring effect of small grade-school classes.[17] Similarly, perhaps, students appear to be willing to stay in school longer if they have been exposed to a more connected grade cohort in middle or secondary school, even absent relative gains in performance.

5. Peer Quality vs. Peer Quantity

To further distinguish traditional peer effects from the effects of connectedness, I explore a setting in which the two effects would work in opposite directions. This could be characterized as a distinction between peer quality and peer quantity. The logic of traditional peer effects is that low quality peers generate low education outcomes for individuals (e.g., because individuals imitate low effort choices of their peers). In addition, low performing peers select peers who resemble them and this selection effect is often confounded with a true peer effect. But having many low-performing peers should if anything produce negative education outcomes in both ways (both because there are more "bad influences" to imitate and because there is a stronger probability that the individual who selected or was selected by these low performers was a low performer to begin with.) The traditional peer effect, then, predicts that having more low-performing

[17] See also Bowles, Gintis, and Osbourne (2001). The paper argues that important effects of schooling, not captured by measures of cognitive skill, are its socialization effects.

peers will be associated with *lower* education outcomes. In contrast, the social utility channel emphasized here is that having more peers (even "bad" ones) raises social utility relative to the outside option and keeps one in school *longer*, the opposite prediction.[18] Similarly, if the results in Table III, Columns 4 and 5, are driven by connectedness or social utility, then having peers *who themselves have more peers (*i.e., being in a more connected environment*)*, raises outcomes, even if one's peers' peers are "bad" peers.

Panel A of Table V restricts the Wave III sample to respondents whose nominated friends had an average GPA higher than the school's average GPA. These students, then, had high-performing peers. Column 1 indicates, as expected, that having more friends of this type increased one's own academic performance. Panel B restricts the sample to individuals with low-performing peers. Column 1 indicates that having more friends of this type was associated with having earned a (marginally) lower GPA. Again, this is the expected peer effect (or the result of selection by low-performers into low-performing peer groups.)

Panel B, Columns 2 and 3, however, flesh out the story in a provocative way. Even if one's peers are **low performers**, having more of them is strongly associated with a **higher probability of finishing high school and going on to college**. (The same story holds when the dependent variable is years of education.) The evidence would suggest that the connectedness effect dominates the traditional peer effect. Moreover, as mentioned above, it is likely that low-performing students select low-performing friends. Despite the effects of selection, students with more low-performing friends stay in school

[18] For the above interpretation to hold, it would need to be the case that social ability had a relatively minor effect on utility in the workplace, but a large effect on utility in school.

longer. Evidence suggests that the effects of connectedness dominate **both** the standard peer effect and the selection effect.

GPA, however, may be an inadequate characterization of peer quality. It is not possible to observe long-run outcomes for most peers in the network (as only a subset are in the Wave III sample), but it is possible to observe self-reported education expectations for respondents in the large in-school sample, and to characterize "good" peers as those who expect to finish college. Panels C and D repeat the exercise above using this revised characterization of peer quality, and the pattern of the results is quite similar Students with high quality peers are defined as those for whom more than half of their connections expect to graduate from college. In Panel D, for example, the individuals in the sample have friends who do not expect to graduate from college, on average. Column 1 indicates that having more peers of this type is associated with having earned a lower GPA. Despite this, Columns 2 and 3 indicate that when one's peers are more numerous, one's own education attainment **rises.**

Having more connections may indicate the individual possesses positive traits associated with staying in school longer. Therefore, as a final test, the analysis in Panels E through H relies not on the number of friends the individual has, but on the average number of her friends' friends. The finding is that if the friends of the individual's friends are more numerous, then the individual experiences higher education outcomes, even if her friends' friends are low performers. [19] [20]

---

[19] This may also help distinguish between the two connectedness channels discussed in section 2: social utility and knowledge spillovers. Knowledge spillovers resemble the standard peer effect: If knowledge spillovers drive the observed effects of connectedness on educational attainment, then one would expect that students in connected settings exhibit would more "knowledge," as measured by GPA. However, for students with low-performing peers (or peers' peers), having more connected peers is associated with **lower GPA** but higher educational attainment. Though not definitive, this would seem to argue against the knowledge-spillover channel, at least for this low-performing subsample.

6. Neighborhood Friends

School friends and neighborhood friends need not be identical, as a given school draws students from a multiplicity of neighborhoods. The Add Health survey data contain only school friends. If neighborhood characteristics influencing long-run outcomes are correlated with an individual's propensity to form neighborhood ties (as opposed to school friendship ties), then the coefficients on connectedness in Table III could be biased upwards. A story about disadvantaged students being bused to distant schools captures the intuition. If disadvantaged students make fewer friends at school because they live far away from their schoolmates, then the number of friendship ties could be a proxy for unobserved disadvantage. The Add Health data contain information on spatial relationships between respondents. Figure III shows the relative locations of residences for students attending School 77. From this data, I computed the average distance between each student and her schoolmates. If distance captures in part the student's propensity to form in-school friendships as opposed to neighborhood ties, and if there exists a systematic correlation between distance and disadvantage, then including distance should lower the coefficient on individual connections. Regressions in Tables II and III already include the distance measure and the coefficient is negative in all specifications.

---

[20] This section also tests the third prediction of Section II: Factors that hinder the production of academic rewards but facilitate the production of social rewards will not generally influence the decision to stay in school for types with low academic ability. The intuition is that agents with low academic ability would have been consuming more social rewards and fewer academic rewards in the first place. An increase in the effort cost of academic rewards does little harm, whereas a decrease in the effort cost of social rewards benefits them. In spite of (marginally) lower GPAs, agents with more low-performing friends consume more education, because social rewards are the dominant component of utility for low-performers.

7. Mobility

Students move between schools, and students new to a school may differ from their schoolmates. In particular, one could imagine that students new to a school make fewer friends and have lower education attainment. To account for the possibility that effects of connectedness could be confounded with effects of new students, the regressions in Tables II and III include an individual "new student" control and a control for the fraction of new students in the school or grade. In Table III, grades with a larger fraction of new students are associated with lower education attainment. The effect of connectedness on education outcomes, however, is robust to the inclusion of these controls.

8. Grade-Specific Teacher Quality

In addition to sharing grade-level peers, individuals in a grade cohort share a common set of teachers. Perhaps talented teachers in a grade-year cohort cause students both to make more friends (or to fill in more names on the questionnaire) and to have higher educational attainment. In a world in which students and teachers did not move between schools, each student would eventually receive the "teacher quality" treatment for each grade in the school. Thus, this channel would seem to be driven primarily by movements of students and/or teachers in and out of schools or by changes in teacher quality over time. One may infer that between Wave I and Wave II, 12% of students in the sample (or less) left their school without having completed the grades offered in that school.[21] Unfortunately, there is no data available on teacher movements and grade-specific teacher quality.

Grade-specific unobservables cannot be ruled out as a possible confounding factor. Arguably, though, the teacher quality explanation, if true, would also suggest that social utility or connectedness (as facilitated by teachers) drives outcomes: "Teacher quality," here, would seem to cause students to claim more friends and to stay in school longer, but to do so without increasing their test scores or their GPAs.

C. Discussion

Comparisons with existing research provide a sense of the magnitude of the coefficient on connectedness. A causal interpretation of the preferred Table III regressions implies that an addition of about 2 links to the average connectedness of grade-level peers would eliminate the college enrollment gap between blacks and whites.[22] Related work suggests that this is not an implausibly large estimate. Ehrenberg and Rothstein(1994) find that students at historically black colleges were 9% to 29% more likely to graduate than similarly qualified black students at other 4-year colleges. And Constantine(1995) finds that they went on to earn higher wages.[23] (Moreover, the findings obtain despite lower average peer SAT scores at historically black colleges.) Though connectedness, as such, is not measured in these studies, one could infer that

[21] This was based on attrition from the sample between Wave I and Wave II other than through finishing the final grade offered in a school, and equates to about 20% of the students who were in non-final grades. It may overstate student movement because other factors could have led to attrition. (Few students changed from one surveyed school to another surveyed school between the two surveys.)
[22] See Maxwell[1994].
[23] The estimated wage gap ranged from 11% to 38%.

connectedness, cultural cohesiveness, or social utility at historically black colleges drives this seemingly large observed effect.

Other evidence of large effects of connectedness on long-run outcomes comes from Galleoti and Mueller (2005), who use the Washington Longitudinal Study to analyze friendship relations in high schools as predictors of adult wages. After controlling for school-specific effects, cognitive ability, grade rank in class, and personality traits, they find that differences in friendship ties in high school predict large wage differences, 35 years later. Galleoti and Mueller's study is related to the present analysis, but differs in important ways. There is only one grade and cohort in the sample—and the students were in high school in 1957. A maximum of 3 friends could be nominated. Information on the full school or grade network was unavailable. Main findings pertained to "isolates" (students with no friends.) Though an advantage of the WLS is that it contains long-run outcomes 35 years after high school, the dataset would seem to limit potential strategies to distinguish the effect of individual endowments from the effect of institutional factors. Moreover, much has changed in 50 years—including the dramatic rise in the wage premium for educated workers.

V. DETERMINANTS OF CONNECTEDNESS

The motivation for exploring a possible relationship between connectedness and long-run outcomes is immediately clear if connectedness is driven in part by policy. It is worth pausing to speculate, given evidence from the data, about institutional factors that could be determinants of connectedness.

Much existing research work suggests that racial heterogeneity influences connectedness.[24] Figure IV shows a scatterplot of number of friendship ties and school racial heterogeneity. A clear negative correlation is visible. (To contrast different types of schools close up, compare the directed graphs in Figures II.A and II.B. The school in Figure II.A—high in heterogeneity and low in connectedness—resides in the lower right portion of the scatterplot in Figure IV, whereas the school in Figure II.B is from the upper left portion.) Do these findings survive the inclusion of controls?

Column 1 of Table VI shows results of an OLS regression of individual connectedness on individual, school, and neighborhood covariates, including school-level racial heterogeneity, H_j: School heterogeneity is negatively correlated with connectedness and significant at the 5 percent level. A high concentration of males appears bad for connectedness, whereas high parental education is associated with increased connectedness. Interestingly, school racial heterogeneity appears not to be a proxy for parental endowment heterogeneity or social class heterogeneity. Standard deviation of parental education (at the school level) appears to be marginally positively correlated with individual connectedness.

A main concern is that schools with large white populations differ in their response to increased heterogeneity from schools with large black populations. Table VI, Columns 2 and 3 summarize regressions on schools in which black students and white students, respectively, represent over 40% of the student population (and are the largest racial group). Both estimated heterogeneity coefficients are negative and significant. It does not appear that homogeneity is a proxy for an increased proportion of white students

[24] See Alesina and La Ferrara (2004) for a detailed survey of existing evidence relating ethnolinguistic fractionalization to measures of trust.

(or for unobserved advantage that might be associated with this.) In Column 4, the negative association between heterogeneity and connectedness is robust to the inclusion of proportion Black, proportion Asian, proportion American Indian, and proportion Hispanic as regressors. To account for neighborhood and school-specific effects, I include school dummies in the model summarized in Column 5 and analyze individual connectedness as a function of racial heterogeneity at the grade level. Grade cohort racial heterogeneity is negatively correlated with connectedness and significant at the 8% level. Being part of a more racially heterogeneous grade cohort within a given school is associated with having fewer friends. [25] [26]

A second potential institutional determinant of connectedness worth mentioning is school size. One rationale behind the recent push for small schools is that smaller schools are more "personal" and allow students to connect more strongly to teachers and classmates.[27] While it is difficult to measure connectedness to teachers, it is possible to study whether students connect more to each other in smaller schools. There is no evidence of this in Table VI. In Columns 1 and 4, the only specifications in which the estimated coefficients on school size are negative, they are not statistically significant—and the point estimates are very small. (In column 4, for example, a reduction of 400

[25] Racial heterogeneity of the grade cohort might seem a plausible instrument for connectedness, given the association between connectedness and heterogeneity in Table V. However, the Wave III sample is much smaller than the Wave I in-school sample. The coefficient on racial heterogeneity is imprecisely estimated in the smaller sample. Thus, well-known problems associated with weak instruments argue against using racial heterogeneity as an instrument for connectedness.
[26] Recent researchers in sociology have also used the Add Health dataset to analyze friendship networks and race (e.g., Moody, 2001). Unlike the present investigation, the primary purpose of these studies has been to describe and understand the determinants of cross-race friendships relative to same-race friendships. There is also research on race and friendship relations in the economics literature. Fryer and Torrelli (2005) focus on same-race friendships. They offer evidence from Add Health friendship data of large racial differences in the relationship between popularity and GPA.
[27] Linda Shaw, The Seattle Times, November 5, 2006.

students in school size is associated with an increase of only one tenth of a link.) Smaller schools do not appear to foster more friendships among students.

VI. SUMMARY AND CONCLUSION

The evidence indicates being part of a more connected cohort within a given secondary or middle school is associated with significantly higher levels of schooling attained and higher probabilities of having attended college, 7 years later. The analysis produced no evidence that smaller schools increase connectedness of the student population, contrary to some educators' assumptions. However, being part of a more racially heterogeneous cohort within a given school is associated with decreased connectedness. This could suggest that some policies designed to harness peer effects may have unintended consequences. Collection and analysis of network data in a controlled setting (with random school or classroom assignment) may be warranted. More generally, it is possible that the collection of network-type data in college or labor market settings[28], and the subsequent application of the economist's statistical toolkit to this data, could yield new and useful insights. These are subjects for future research.

[28] For a prominent example, see Marmaros and Sacerdote(2004), an investigation of social networks in college as determinants of labor market outcomes.

References

Akerlof, George and Rachel Kranton, "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, XL(2002), 1167-1201.

Alesina, Alberto, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, CXIX(1999), 1243-1284.

Alesina, Alberto, and Eliana La Ferrara, “Ethnic Diversity and Economic Performance,” NBER Working Paper 10313, 2004.

Angrist, Joshua and Kevin Lang, "How Important are Peer Effects? Evidence from Boston's Metco Program," NBER Working Paper #9263, 2002.

Betts, Julian, Andrew Zau, and Lorien Rice, *Determinants of Student Achievement: New Evidence from San Diego* (San Francisco: Public Policy Institute of California, 2003).

Borgatti, Steve, Martin Everett, and Linton Freeman, *Ucinet 6.87 for Windows: Software for Social Network Analysis* (Harvard, MA: Analytic Technologies, 2002).

Bowles, Samuel, Herbert Gintis, and Melissa Osbourne, “Incentive-Enhancing Preferences: Personality, Behavior, and Earnings,” *American Economic Review*, XCI(2001), pp.155-158.

Constantine, Jill, “The Effect of Attending Historically Black Colleges and Universities on Future Wages of Black Students,” *Industrial and Labor Relations Review*, XLVIII(1995), 531-546.

Ehrenberg , Ronald and Donna Rothstein, “Do Historically Black Institutions of Higher Education Confer Unique Advantages on Black Students? An Initial Analysis,” in *Choices and Consequences: Contemporary Policy Issues in Education,* R. Ehrenberg, ed. (Ithaca, N.Y.: ILR Press, 1994).

Fryer, Roland and Paul Torelli, “An Empirical Analysis of ‘Acting White’,” NBER Working Paper 11334, 2005

Galeotti, Andrea and Gerrit Mueller, “Friendship Relations in the School Class and Adult Economic Attainment,” IZA Discussion Paper No. 1682, 2005.

Hanushek, Eric, John Kain, Jacob Markman, and Steven Rivkin, "Does Peer Ability Affect Student Achievement?" NBER Working Paper 8502, 2001.

Hanushek, Eric, John Kain and Steven Rivkin, "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement," mimeo, Stanford, 2004.

Hoxby, Caroline, "The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation," *Quarterly Journal of Economics*, CXV(2000), 1239-1285.

Marmaros, David, and Bruce Sacerdote (2002), "Peer and Social Networks in Job Search," *European Economic Review*, XLVI, 870–879.

Manski, Charles, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, LX(1993), 531-542.

Maxwell, Nan, "The Effect of Black-White Wage Differences of Differences in the Quantity and Quality of Education," *Industrial and Labor Relations Review*, XLVII(1994), 249-264.

Moody, James, "Race, School Integration, and Friendship Segregation in America," American Journal of Sociology, CVII (2001), 679-716.

Shaw Linda, "Foundation's Small-schools Experiment Has Yet to Yield Big Results," *The Seattle Times*, November 5, 2006.

Wasserman, Stanley and Katherine Faust (1994), <u>Social Networks Analysis: Methods and Applications,</u> Cambridge: Cambridge University Press.

Table I
Descriptive Statistics

	Mean	St. Dev.
Links	8.74	5.68
Out Degree	4.37	3.05
In Degree	4.37	3.70
Age	14.85	1.77
Parental Ed	13.05	2.22
Parental Incom (1000s)[a]	35.84	44.62
AH Vocab Test[a]	97.62	25.52
Distance (ave. km, peers)[a]	7.12	6.54
GPA	2.44	1.21
New Student	0.24	0.43
White	0.58	0.49
Black	0.21	0.41
Asian	0.04	0.19
Amer. Ind.	0.04	0.19
Hisp	0.10	0.30
Male	0.51	0.50
Grade 7	0.17	0.38
Grade 8	0.16	0.37
Grade 9	0.17	0.38
Grade 10	0.17	0.37
Grade 11	0.15	0.36
Grade 12	0.16	0.37
Het. Race (sch)	0.39	0.19
Par. Ed. (sch)	13.00	0.82
Std. Dev. Par. Ed. (sch)	2.07	0.26
Class Size(sch)	25.27	4.79
Size (sch)	938.22	664.15
Prop. New Teachers(sch)	9.89	14.88
Prop. PTA Participation(sch)	24.37	22.68
GPA(sch)	2.86	0.25
Prop. New Students(sch)	0.26	0.13
Prop. Entry Grade	0.19	0.39
Prop. Male(sch)	0.51	0.06
Juv. Crime (per 100K, cnty)	339.32	172.81
Het. Race (cnty)	0.28	0.18
Med. Income (cnty: in 1000s)	28828.54	7871.16
Std. Dev. Income (cnty)	28675.23	5421.68
Min Wage (state)	4.28	0.16
Crime Rate (per k, state)	5184.90	1105.18

[a]These variables available for In-home sample. All other covariates available for full sample.

Table II
Connectedness and Long Run Outcomes[a,b]

Dependent Variable	Yrs Schooling (OLS) 1	College (Probit) 2	Employed (Probit) 3	Yrs Schooling (OLS) 4	College (Probit) 5	Employed (Probit) 6
Links (ind)	.0423***	.0344***	.0162***	-	-	-
	(.00423)	(.00361)	(.00538)			
		[.0099]	[.0047]			
Links (sch)	-	-	-	.201***	.132***	.0966***
				(.0469)	(.0458)	(.0361)
					[.0386]	[.0279]
Parental. Education (ind)	.127***	.109***	.00878	.133***	.113***	.0109
	(.0123)	(.0104)	(.0135)	(.0127)	(.0106)	(.0135)
Family Income: 1000s (ind)	.00193***	.00423***	.000719	.00227***	.00458***	.000831
	(.000637)	(.00111)	(.000909)	(.000681)	(.00119)	(.000921)
GPA(ind)	.69***	.539***	.0689**	.724***	.561***	.0772**
	(.0407)	(.0305)	(.0323)	(.0414)	(.0305)	(.0319)
AH Vocab (ind)	.0143***	.0132***	.00376**	.015***	.0136***	.00405**
	(.00161)	(.00173)	(.0018)	(.00168)	(.00175)	(.00182)
New Student(ind)	-.0373	.0291	-.0147	-.1	-.0254	-.0389
	(.0707)	(.0702)	(.0715)	(.0704)	(.0689)	(.0726)
Distance (ind)	.00038	-.00369	-.00036	-.00112	-.00466	-.00117
	(.0039)	(.00295)	(.00362)	(.00382)	(.00285)	(.00352)
Ave. Parent Ed. (sch)	.454***	.395***	.104*	.429***	.372***	.0914
	(.0575)	(.0572)	(.0576)	(.0576)	(.0559)	(.0578)
Std. Dev. Parent Ed. (sch)	.00381	-.00476	.126	-.0446	-.0422	.0876
	(.215)	(.178)	(.153)	(.22)	(.174)	(.155)
Het. Race (sch)	-.375	-.232	-.308*	-.245	-.151	-.231
	(.251)	(.212)	(.177)	(.246)	(.212)	(.184)
Ave. GPA (sch)	-.415***	-.446***	-.0946	-.477***	-.473***	-.126
	(.151)	(.129)	(.115)	(.144)	(.13)	(.11)
Prop. Male(sch)	.322	-.395	.134	.972	-.0209	.502
	(.793)	(.563)	(.925)	(.795)	(.578)	(.928)
Prop. New Student (sch)	.152	.0367	.0911	.282	.114	.151
	(.219)	(.229)	(.235)	(.214)	(.227)	(.23)
Size: 100s (sch)	.00294	-.00411	.000631	.00584	-.00229	.00211
	(.00761)	(.00627)	(.00493)	(.00728)	(.00607)	(.00505)
Income: 1000s (cnty)	-.00701	-.0128	.0242***	-.0086	-.0139*	.0229**
	(.00876)	(.00808)	(.00863)	(.00865)	(.00799)	(.00879)
Het(cnty)	.0922	.125	.0751	-.0388	.048	.0102
	(.313)	(.341)	(.292)	(.323)	(.346)	(.282)
Min wage(state)	.0584	-.0972	.0515	-.0171	-.148	.0143
	(.148)	(.154)	(.18)	(.154)	(.155)	(.181)
Obs	10141	10141	7121	10141	10141	7121

[a]Standard errors in parentheses. Regressions include the full set of covariates from Table I.
[b]Weighted population mean of marginal effects in brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table III
Connectedness and Long Run Outcomes - School Fixed Effects[a,b]

Dependent Variable	Yrs School (OLS)	College (Probit)	Employed (Probit)	Yrs School (OLS)	College (Probit)	Employed (Probit)	Yrs School (IV)	College (IV-Prob)	Employed (IV-Prob)
	1	2	3	4	5	6	7	8	9
Links (ind)	.0393***	.0337***	.0136**	-	-	-	.187***	.0819*	-.0151
	(.00432)	(.00363)	(.00578)				(.0434)	(.0419)	(.0437)
		[.0094]	[.0038]					[.0232]	[-.0042]
Links (grd)	-	-	-	.12***	.0607**	-.01	-	-	-
				(.0279)	(.0278)	(.0303)			
					[.0172]	[-.0028]			
Parental Ed (ind)	.126***	.111***	.00829	.133***	.115***	.0103	.102***	.0128	.0128
	(.0119)	(.011)	(.0137)	(.0125)	(.0113)	(.0137)	(.0139)	(.0152)	(.0152)
Fam. Inc: 1000s (ind)	.00204***	.00451***	.000837	.00241***	.00488***	.00095	.000827	.00104	.00104
	(.000621)	(.00118)	(.000914)	(.000664)	(.00127)	(.000928)	(.000714)	(.000928)	(.000928)
GPA(ind)	.716***	.568***	.0882**	.747***	.589***	.0959***	.604***	.106**	.106**
	(.0403)	(.0322)	(.034)	(.0411)	(.0321)	(.0336)	(.0572)	(.0453)	(.0453)
AH Vocab (ind)	.013***	.0127***	.00376*	.0138***	.0132***	.00399**	.0102***	.00414**	.00414**
	(.00152)	(.00178)	(.00194)	(.00158)	(.00179)	(.00197)	(.00192)	(.0019)	(.0019)
New Student (ind)	.0642	.0523	.0307	-.0161	-.0205	.00739	.345***	-.0241	-.0241
	(.0784)	(.08)	(.0892)	(.0784)	(.0805)	(.0896)	(.107)	(.122)	(.122)
Distance (ind)	-.00626*	-.00998***	-.000974	-.00811**	-.0111***	-.00146	-.000445	-.0022	-.0022
	(.00364)	(.00341)	(.00438)	(.00373)	(.00347)	(.00424)	(.00455)	(.00444)	(.00444)
Ave. Parent Ed. (grd)	-.0628	.0357	.0478	-.136	.000192	.0531	-.0849	.0457	.0457
	(.119)	(.114)	(.127)	(.12)	(.115)	(.131)	(.138)	(.128)	(.128)
S.D. Parent Ed. (grd)	.362**	.211	-.242	.384**	.206	-.253	.415**	-.258	-.258
	(.179)	(.163)	(.161)	(.175)	(.162)	(.16)	(.166)	(.161)	(.161)
Het. Race (grd)	.158	.308	-.471	.419	.424	-.504	.375	-.478	-.478
	(.759)	(.609)	(.75)	(.723)	(.62)	(.754)	(.738)	(.751)	(.751)
Ave. GPA (grd)	-.317**	-.149	-.178	-.32**	-.157	-.189	-.224	-.196	-.196
	(.148)	(.121)	(.176)	(.143)	(.119)	(.176)	(.155)	(.178)	(.178)
Male(grd)	.0425	-.0628	-.184	.091	-.0634	-.196	.178	-.185	-.185
	(.4)	(.335)	(.36)	(.371)	(.328)	(.364)	(.405)	(.364)	(.364)
New Student (grd)	-.362***	-.219*	-.147	-.265**	-.13	-.132	-.631***	-.104	-.104
	(.118)	(.125)	(.141)	(.118)	(.126)	(.14)	(.142)	(.159)	(.159)
Size: 100s (grd)	-.0262	-.104	.0796	-.0706	-.12	.0847	-.0698	.0834	.0834
	(.0923)	(.0805)	(.0584)	(.0872)	(.0758)	(.061)	(.0928)	(.0591)	(.0591)
Obs	10141	10141	7121	10141	10141	7121	10141	10141	7121

[a]Standard errors in parentheses. Regressions include full set of covariates from Table I (except where collinear by definition).
[b]Weighted population mean of marginal effects in brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table IV
Connectedness and Long Run Outcomes - Robustness Checks

	Dependent Variable:					
	Years School	**Years School** (Sch Size>100)	**Years School** (Sch Size>250)	**College**	**College** (Sch Size>100)	**College** (Sch Size>250)
	(OLS)	(OLS)	(OLS)	(Probit)	(Probit)	(Probit)
	1	2	3	4	5	6
Connectedness Measure:						
Links	.0393***	.0395***	.0406***	.0337***	.0332***	.0331***
	(.00432)	(.00446)	(.00454)	(.00363)	(.00372)	(.00376)
Out Degree	.054***	.0546***	.0574***	.0449***	.0454***	.0469***
	(.00736)	(.00761)	(.00759)	(.0073)	(.00743)	(.00746)
In-Degree	.0506***	.051***	.0515***	.044***	.043***	.0419***
	(.00582)	(.006)	(.0062)	(.00492)	(.00498)	(.005)
Links (grd)	.12***	.116***	.114***	.0607**	.0625**	.0628*
	(.0279)	(.0283)	(.0323)	(.0278)	(.0281)	(.0324)
Out Degree (grd)	.173***	.172***	.17***	.091**	.0956**	.0907
	(.0408)	(.0431)	(.053)	(.0403)	(.0436)	(.0551)
In-Degree (grd)	.186***	.182***	.169***	.0846*	.0825*	.0619
	(.0439)	(.0466)	(.0546)	(.0452)	(.0461)	(.0555)

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parentheses
All regressions include school dummies and full set of controls from Table III.
Each row-column entry (with associated standard error beneath it) represents a separate regression.

Table V

Connections, GPA, and Education Attainment[a]

Dependent Variable:	GPA (OLS) 1	Grad HS (Probit) 2	College (Probit) 3
A. Ave. GPA (Peers) >= Ave. GPA (School)			
Links (ind)	.0185***	.026***	.0484***
	(.00219)	(.00934)	(.00689)
Ave. GPA (peers)	.642***	.0725	.476***
	(.0702)	(.328)	(.111)
Obs	4423	4423	4423
B. Ave. GPA (Peers) < Ave. GPA (School)			
Links (ind)	-.00257	.0261***	.0209***
	(.00308)	(.00606)	(.00671)
Ave. GPA (peers)	.235***	.303***	.349***
	(.079)	(.106)	(.105)
Obs	4372	4372	4372
C. (Fract. Peers Expect to Finish College)>.5			
Links (ind)	.00929***	.0194***	.033***
	(.00166)	(.00589)	(.00605)
Ave. GPA (peers)	.613***	.401***	.734***
	(.0418)	(.106)	(.0517)
Obs	6577	6577	6577
D. (Fract. Peers Expect to Finish College)<=.5			
Links (ind)	-.0127**	.0474***	.0303**
	(.00601)	(.0142)	(.0127)
Ave. GPA (peers)	.296***	.499***	.353***
	(.0901)	(.138)	(.131)
Obs	1142	1142	1142

Dependent Variable:	GPA (OLS) 1	Grad HS (Probit) 2	College (Probit) 3
E. Ave. GPA (Peers' Peers) >= Ave. GPA (School)			
Links (peers)	-.0106***	.0122	.0463***
	(.00389)	(.0166)	(.0087)
Ave. GPA (peers)	1.14***	.754***	1.02***
	(.0625)	(.205)	(.217)
Obs	4096	4096	4096
F. Ave. GPA (Peers' Peers) < Ave. GPA (School)			
Links (peers)	-.0105**	.0193**	.0183**
	(.0045)	(.0094)	(.00892)
Ave. GPA (peers)	1.23***	1.02***	.903***
	(.0868)	(.176)	(.139)
Obs	3797	3797	3797
G. (Fract. Peers' Peers Expect to Finish College)>.5			
Links (peers)	.00527	.019**	.0387***
	(.00346)	(.00944)	(.00782)
Ave. GPA (peers)	-1.31***	-1.37***	-2.37***
	(.131)	(.298)	(.314)
Obs	7553	7553	7553
H. (Fract. Peers' Peers Expect to Finish College)<=.5			
Links (peers)	-.0173	.0701**	.0323
	(.0106)	(.0297)	(.0291)
Ave. GPA (peers)	.305	.285	-.823
	(.416)	(.695)	(.857)
Obs	331	331	331

[a]Standard errors in parentheses. Regressions include the full set of covariates from Table I.

* significant at 10%; ** significant at 5%; *** significant at 1%

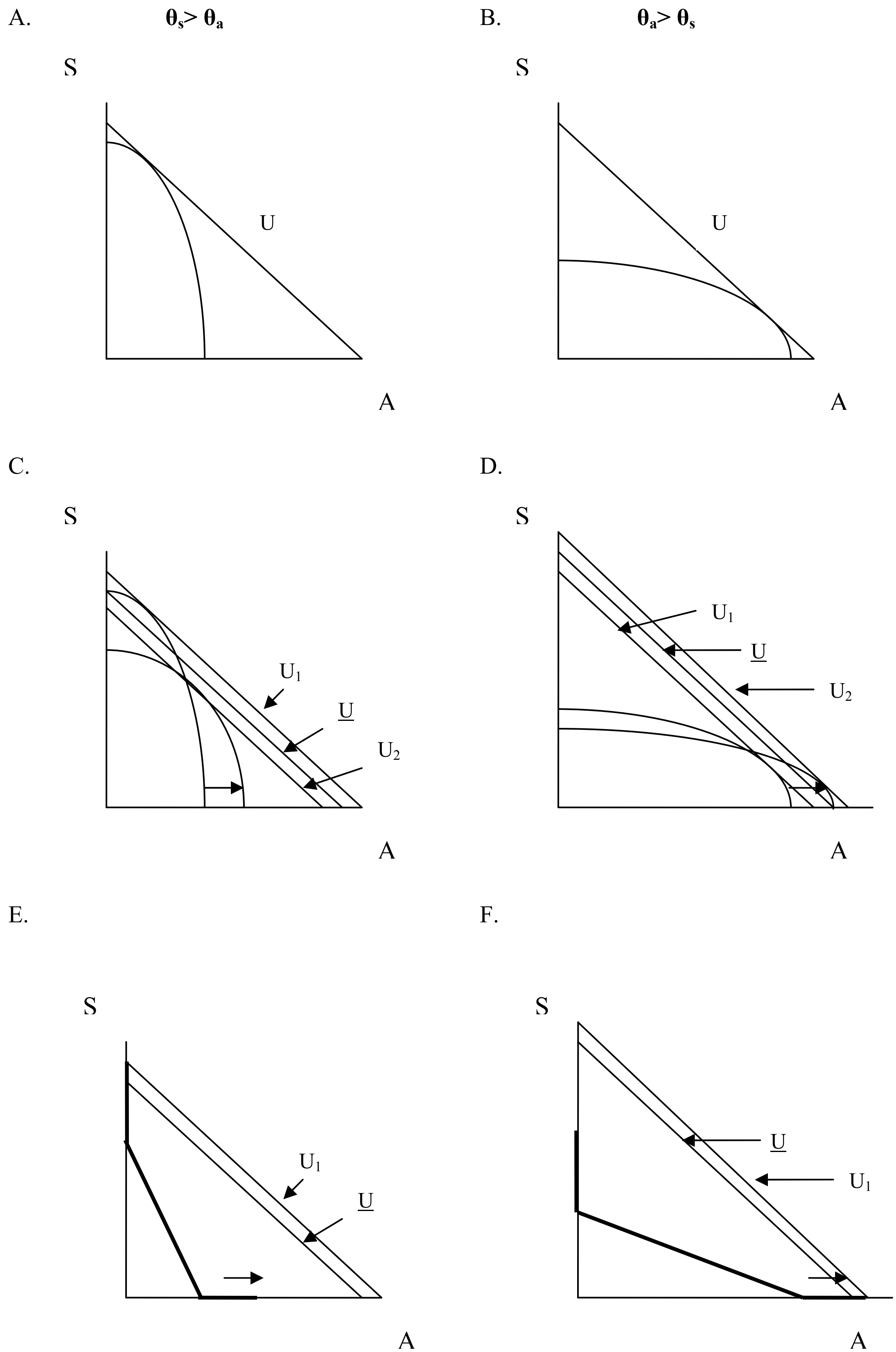
Peers who "expect to finish college" are respondents who rate their chances at better than "50-50", at the time of the in-school survey.

Table VI
Determinants of Connectedness[a]

Dependent Variable	**All** (Links) 1	**Prop. Black>.4** (Links) 2	**Prop. White>.4** (Links) 3	**All** (Links) 4	**School Fixed Eff.** (Links) 5
Het. Race (sch or grd)[c]	-2.07**	-5.23**	-4.67***	-1.86**	-3.14*
	(.816)	(2.17)	(1.42)	(.856)	(1.76)
Prop. Black (sch or grd)[c]	-	-	-	-1.77***	.79
				(.639)	(1.9)
Prop. Asian(sch or grd)[c]	-	-	-	-3.47*	6.3***
				(2.04)	(2.4)
Prop. Amer. Ind. (sch or grd)[c]	-	-	-	-2.63	3.46
				(5.7)	(4.27)
Prop. Hisp (sch or grd)[c]	-	-	-	-.825	-2.06
				(1.42)	(2.55)
Size (100s - sch or grd)[c]	-.0211	.196***	.0139	-.0245	.439**
	(.0227)	(.0464)	(.0268)	(.0225)	(.175)
Prop. Male (sch or grd)[c]	-7.19***	-6.46	-8.19***	-8.13***	-.887
	(1.68)	(3.88)	(2.51)	(1.54)	(.612)
Ave GPA (sch or grd)[c]	.036	4.28***	-.963	-.149	-.774*
	(.558)	(1.15)	(.667)	(.553)	(.442)
Ave. Parent Ed. (sch or grd)[c]	-.0339	-1.24**	.0206	.0299	.533
	(.191)	(.478)	(.247)	(.222)	(.342)
Std. Dev. Par. Ed. (sch or grd)[c]	.914	-4.25**	1.19*	.994	.271
	(.607)	(1.57)	(.634)	(.646)	(.481)
Prop. New Stdnt (sch or grd)[c]	-.482	1.67	-1.11	-.41	2.06***
	(1.06)	(1.24)	(1.15)	(.984)	(.502)
Parental. Education (ind)	.208***	.147***	.238***	.207***	.199***
	(.0198)	(.0496)	(.0256)	(.0196)	(.0198)
GPA(ind)	.707***	.398***	.83***	.716***	.733***
	(.0535)	(.0981)	(.0647)	(.0524)	(.0504)
Male(ind)	-.586***	-.586***	-.721***	-.575***	-2.23***
	(.0722)	(.0722)	(.196)	(.0714)	(.12)
New Student(ind)	-2.02***	-1.81***	-2.06***	-2***	-.556***
	(.174)	(.181)	(.228)	(.171)	(.0706)
Het. Race (cnty)	.705	-.255	3.34*	2.26	-
	(1.19)	(3.03)	(1.7)	(1.46)	
Med. Income. (cnty: 1000s)	.0329	.0471	-.00756	.0305	-
	(.0378)	(.143)	(.0386)	(.0375)	
Std. Dev. Income (cnty)	-.131**	-.13	-.0785	-.123**	-
	(.058)	(.174)	(.0606)	(.0587)	
Obs	66162	10508	42481	66162	66162

[a] Regressions include In-school survey covariates from Table I (except where collinear by definition).
* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parentheses
[c] School level covariates used in columns 1,2,3,4,5,6; grade level covariates are used in column 7.

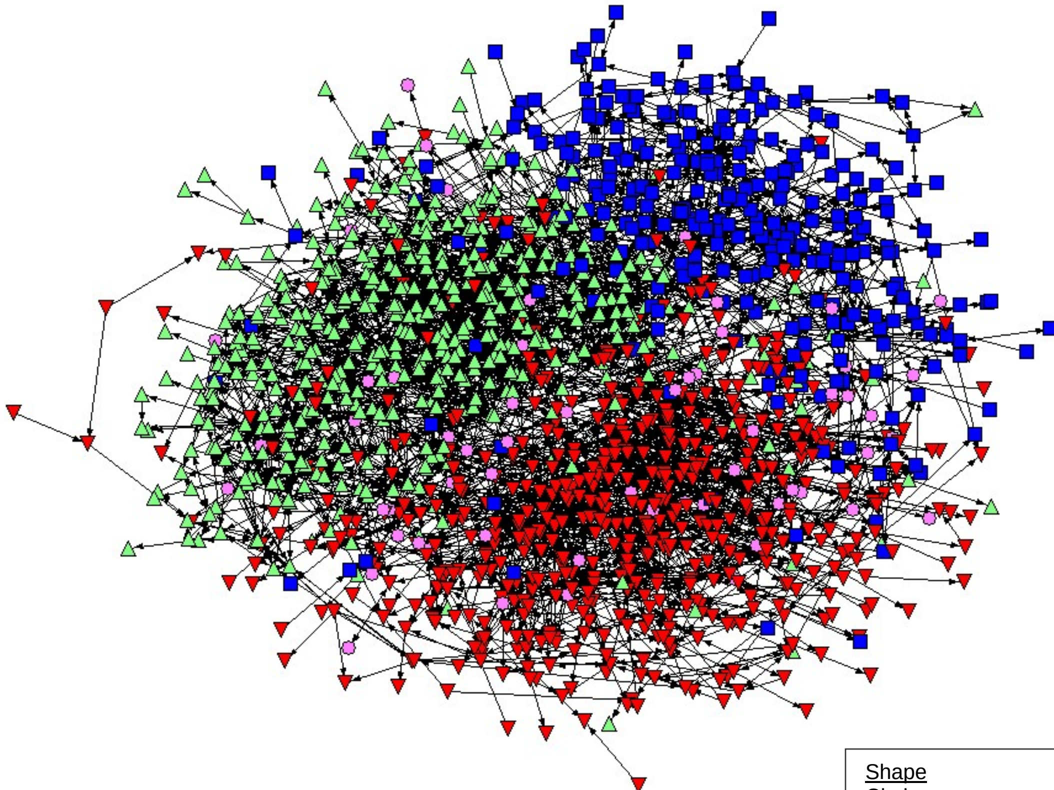
Figure I



A.
$\theta_s > \theta_a$
S
U
A
B.
$\theta_a > \theta_s$
S
U
A
C.
S
U_1
$\underline{U}$
U_2
A
D.
S
U_1
$\underline{U}$
U_2
A
E.
S
U_1
$\underline{U}$
A
F.
S
$\underline{U}$
U_1
A

Figure II

A. Het=.68 Links=5.9



Shape	Race
Circle	White
Square	Black
Up Triangle	Asian
Dwn Triangle	Hispanic

B. Het=.37 Links=8.74

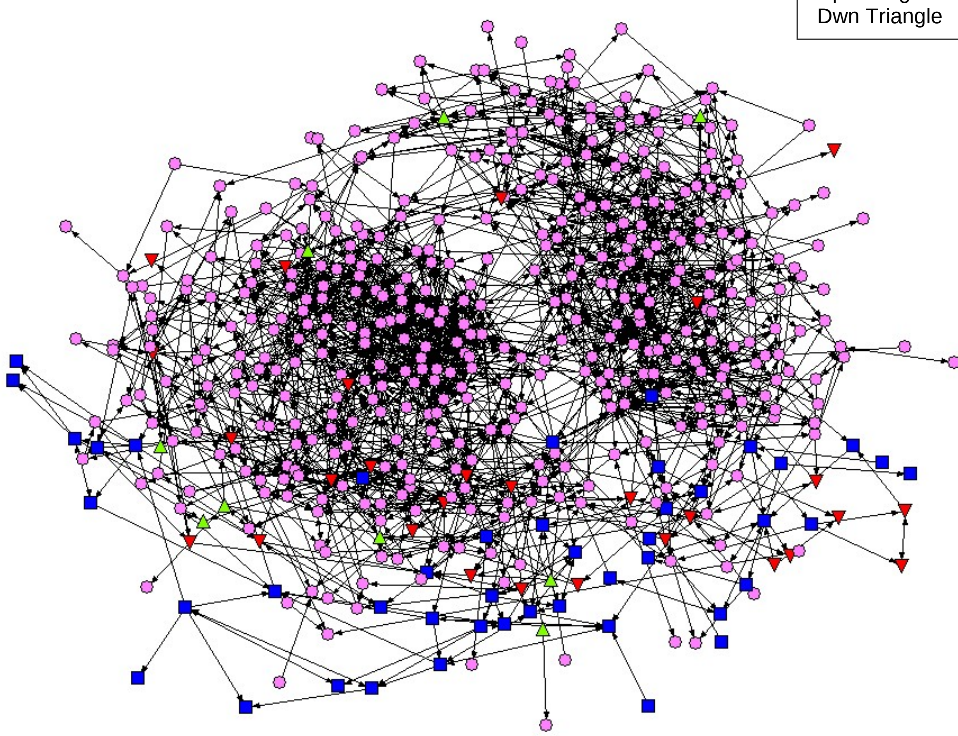
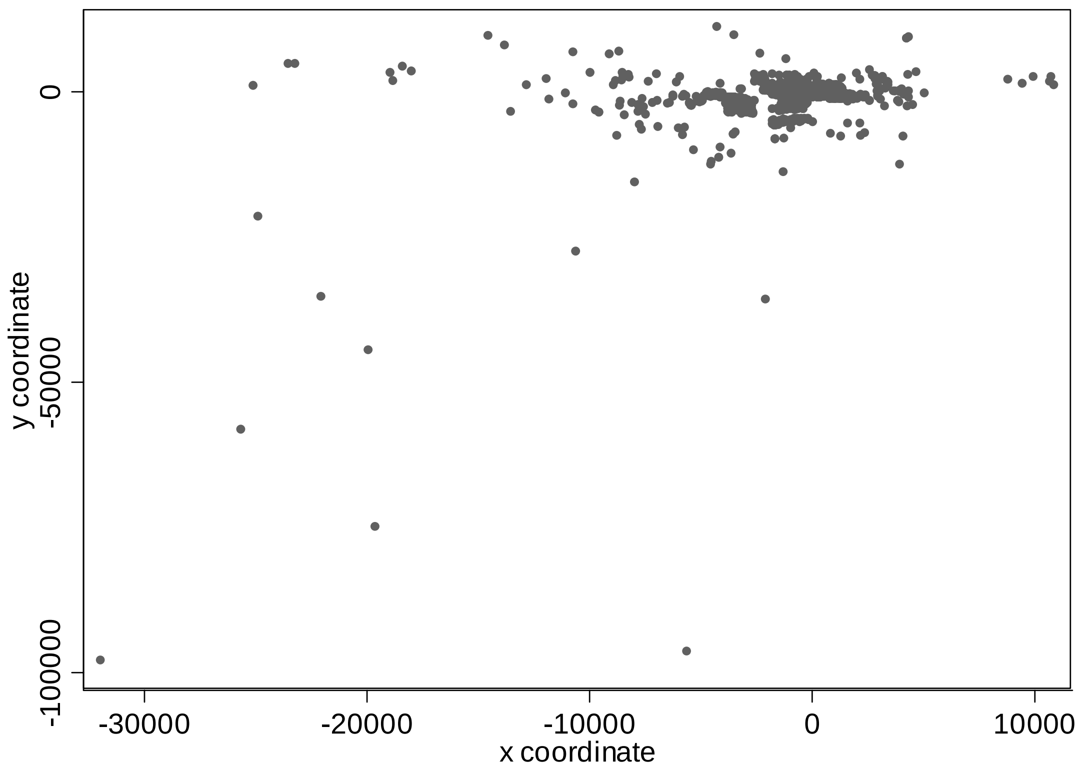


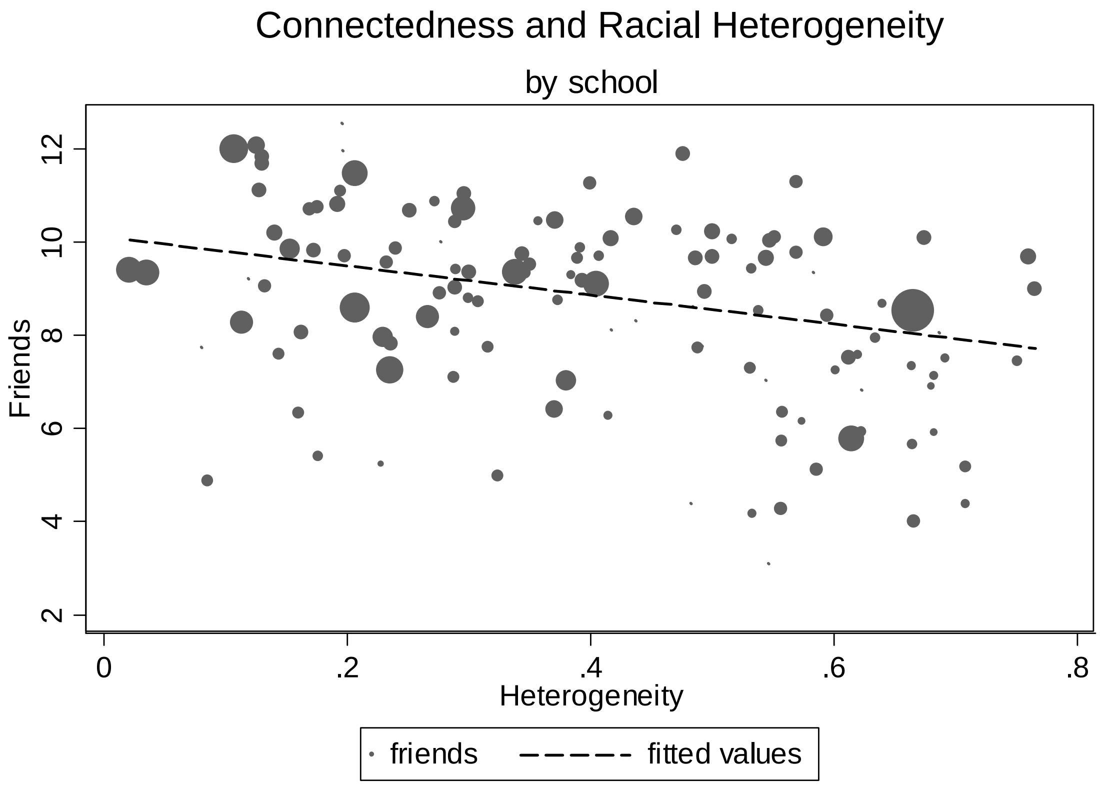
Figure III

School Residences*



*x, y coordinates in meters

Figure IV


Connectedness and Racial Heterogeneity
by school
Friends
2
4
6
8
10
12
0
.2
.4
.6
.8
Heterogeneity
friends
fitted values