HOW DO FRIENDSHIPS FORM?*

David Marmaros and Bruce Sacerdote

We examine how people form social networks among their peers. We use a unique data set that tells us the volume of email between any two people in the sample. The data are from students and recent graduates of Dartmouth College. First-year students interact with peers in their immediate proximity and form long-term friendships with a subset of these people. This result is consistent with a model in which the expected value of interacting with an unknown person is low (making traveling solely to meet new people unlikely), while the benefits from interacting with the same person repeatedly are high. Geographic proximity and race are greater determinants of social interaction than are common interests, majors, or family background. Two randomly chosen White students interact three times more often than do a Black student and a White student. However, placing the Black and White student in the same freshman dorm increases their frequency of interaction by a factor of three. A traditional "linear in group means" model of peer ability is only a reasonable approximation to the ability of actual peers chosen when we form the groups around all key factors including distance, race and cohort.

I. Introduction

It is frequently argued that friends and peers have a large influence on how we behave, how much education we obtain, what career we pursue, and even whom we marry.[1] Families self-select into certain neighborhoods, and students into certain schools because of the perceived peer effects as in Hoxby [2000b], Winston [1999], and many others. However, less has been written on how we actually choose and are chosen by a specific group of friends within a neighborhood, school, or workplace.[2] We find that long-term friendships grow from chance meetings and that small and random differences in proximity have a big impact on our circle of friends.

One reason for the lack of studies on friendship is the scarcity

* 6106 Rockefeller, Dartmouth College, Hanover, NH 03755. Telephone: 603 646 2121. Fax: 603 646 2122. The authors thank the National Science Foundation for generous funding and thank discussants and seminar participants at the Wharton School, DELTA, Ohio State University, and the NBER for helpful suggestions. We thank two anonymous referees and Alberto Alesina, Lawrence Katz, and Edward Glaeser. JEL Classification Codes: I2, J0. Keywords: Social Networks, Peer Effects.

1. See, for example, Harris [1998], Case and Katz [1991], Evans, Oates, and Schwab [1992], Zimmerman [2003], Hoxby [2000a], and Marmaros and Sacerdote [2002].

2. One exception is Falk and Kosfeld [2003] who study the shape of networks in an experimental setting.

The Quarterly Journal of Economics, February 2006

of large micro data sets in which we can identify who is friends with whom, with the notable exceptions of Case and Katz [1991] and Holahan, Wilcox, and Burnam [1978].[3] We solve this problem by measuring the level of social interaction between any two individuals as the volume of email exchanged between the two people during the prior thirteen months. The subjects are students and recent alumni at Dartmouth College. Our exercise is particularly interesting given the random assignment of students to rooms and dorms during their freshman year. The exogenous shock of random assignment allows us to test the power of geographic proximity against other potentially important factors like race, family background, and common interests like athletic teams.

Our methodology also provides a very direct measure of the amount of racial segregation on a campus. Bowen and Bok [1998] explain that most selective universities have made a major push during the last 30 years to increase the racial diversity of their student bodies. However, as argued in Richards [2002], the universities' objectives may be partially blunted if the White and non-White groups on campus spend very little time interacting.

In the recent Supreme Court cases addressing affirmative action, eight universities (Dartmouth College, Harvard, Yale, and Brown Universities, University of Chicago, Duke University, University of Pennsylvania, and Princeton University) jointly filed an amicus brief which emphasized the importance of racial diversity in the educational process. The brief argues that students educate each other and that several studies [Bowen and Bok 1998; Bowen and Levin 2003; Epstein 2002] demonstrate that cross-racial learning takes place and is valued by students and the labor market.

However, other than Duncan, Boisjoly, Kremer, Levy, and Eccles [2003a] few large-scale studies actually measure whether much interracial interaction is taking place. If anything, the evidence we have from campus newspapers and personal anecdotes suggests massive amounts of racial segregation on nearly every campus. See, for example, Schapiro [2003] describing Emory University or Hills [2003] describing Bryn Mawr.

In a test of Bowen and Levin's [2003] thesis, we are able to

3. For example, the NLSY and GSS do ask respondents several questions about their friends, but not enough to allow the sort of detailed analysis we propose here.

ascertain the degree to which athletes or minority students are either isolated from the rest of campus or systematically interacting with peers who have lower academic ability. For example, we show that more than half of a Black student's interactions take place with non-Black students.

We add to the existing literature on friendship or social interactions in several ways. First, we have a more detailed measure of the level of social interaction than has been possible with prior studies. Second, we explore the relative importance (in determining social interactions) of geography, architecture, race, athletic interests, social interests, intellectual interests, and family background. We explore how the importance of these factors varies within- versus across-race and within- versus across-gender relationships. And finally we ask whether a linear in means approach captures the social influences experienced by a student.

Our regressions and simulations lead us to three related conclusions: first, proximity does have a large effect (3×) on the likelihood of social interaction among individuals, regardless of race or family background. However, proximity is not the most powerful policy tool for increasing interracial interactions on a given campus, because the proximity effect is only important over short distances (i.e., within building). Given the physical reality that a student can be housed in truly close proximity to only 45 or so other students, it would be difficult to generate large increases in the total amount of interracial interaction simply through more mixing of the housing.[4]

In contrast, placing two students in the same entering class (cohort) has a 6× effect on the frequency of their interacting, even if the two students are of a different race or are at different ends of the academic ability distribution. Thus, overall cohort composition is important in determining peer group, and this fact can be used to influence the number of interracial interactions or interactions with high SAT scorers experienced by the modal Dartmouth student.

Second, a simple group means or linear in means model of peer influences does not necessarily assign students to their true peer group, as measured by the number of email interactions. The majority of existing peer effects studies use a linear in means

4. For an average White student only about 9 percent of her interactions with Black students involve Black students from her freshman dorm. While proximity increases the likelihood of interaction for any pair of students, same dorm interracial interactions are still a modest fraction of overall interracial interactions.

approach to creating measures of peer background ability or peer outcomes. Researchers typically use the mean outcome or mean pretreatment characteristic for a group which is assumed to represent the individual's peers or friends. For some examples, see Graham [2005], Betts and Zau [2004], Hoxby [2002], and the studies referenced above. The econometrics of social interactions literature including Manski [1993] and Graham and Hahn [2005] uses the linear in means model as a starting point.

A large question for the literature is whether the group means approach approximates the peer influences a student or subject actually experiences. Studies of peer effects at the university level [Sacerdote 2001; Zimmerman 2003; Stinebrickner and Stinebrickner 2004; Duncan et al. 2003b; Foster 2003, 2004; Arcidiacono, Foster, and Kinsler 2003] calculate peer means at the room, hallway, and dorm level. Our data indicate that peer groups constructed in this way can be a reasonable approximation to the true peer groups that form only if we construct peer groups along all dimensions that matter including race, entering class, and geographic distance. And we estimate significantly larger peer effects when we use a more appropriate definition of peer group rather than a simple dorm or hallway mean.

Our third conclusion is that having a minority roommate or dormmate does not appear to lead students into a broader social network of minority students. When a White student is assigned a Black hallmate, she experiences additional interactions with that hallmate but not with other Black students living elsewhere on campus.

Finally, by looking at the same students over time, we discuss how social interactions change following the students' departure from campus after graduation. The panel aspect of the data allow one test of the Gaspar-Glaeser thesis [1998] that email communication is a complement to face-to-face communication, rather than a substitute.

I.A. On Peers, Race, and Location

There is a burgeoning literature on peer effects at the elementary, secondary, and postsecondary levels of education. Hoxby [2000a] finds large peer effects in reading and math test scores among elementary school students. Case and Katz [1991] and Evans, Oates, and Schwab [1992] show that peers are influential in determining risky youth behaviors including drug use, criminal activity, and unprotected sex. A series of papers including Sacerdote [2001], Zimmerman [2003], Stinebrickner and Stinebrickner [2004], Foster [2003], and Duncan, Boisjoly, Kremer, Levy, and Eccles [2002, 2003] use college or university roommates to examine peer effects on both academic and social (particularly drinking) outcomes.

Like us, several authors including Festinger, Schachter, and Back [1963], Abu-Ghazzeh [1999], and Holahan, Wilcox, and Burnam [1978] have emphasized the importance of geographic proximity in determining who interacts with whom. Festinger, Schachter, and Back gathered data on social interactions among new Massachusetts Institute of Technology students in MIT-owned housing. Glaeser and Sacerdote [2000] show that individuals in more dense housing structures are much more likely to interact with their neighbors.

Duncan, Boisjoly, Kremer, Levy, and Eccles [2003] show that the racial composition of freshman housing assignments can have a long-run impact on student attitudes. For example, if student X is randomly assigned a Black roommate, X is somewhat more likely to support affirmative action in admissions and societal income redistribution. We show that housing assignments lead to long-run social interactions among roommates and dormmates both within and across races.

Several psychology researchers have studied the determinants of friendship, and the results of Rainio [1966] and Tuma and Hallinan [1979] imply that similarity and status are two important factors. Waller [1938] and Blau [1964] develop models in which offers of friendship are made and accepted or rejected based on the costs and benefits of the relationship.

I.B. Modeling the Friendship/Peer Group Formation Process

In conducting our analysis, we have in mind a certain model of how friendships form and blossom. Every potential social interaction has associated costs and benefits. The benefits are both a) a flow of information and ideas and b) the utility from sharing a common experience and conversation with another human being. The utility from the common experience component is assumed to increase with the number of previous social interactions that one has had with this specific person. The costs are the time it takes to have the face-to-face conversation, phone conversation, or email exchange. Perhaps the biggest time cost of all is finding out that the other person exists and might be a useful person with whom to speak.[5]

Distance presents itself as a big cost when the value of the social interaction is unknown and especially if the person with whom the interaction might take place is unknown. Common background, interests, and race between two people could raise or lower the benefits of a given social interaction. For example, a White senior from Newton, MA, may have little in common with a Black freshman from Chicago. This might increase the benefits of the interaction to both since the two people have disjoint sets of information. On the other hand, if the goals and concerns of the two people are also completely orthogonal, then the value of the interaction may be low despite a large knowledge gap between the two.

With some functional form assumptions, we might write each agent's expected utility from a potential interaction as

$$E[U(\cdot)] = E[f(\text{information gathered}) + g(\text{shared experience benefit})] - c(\text{time used}),$$

where $g(\cdot)$ is a function that increases with the number of previous social interactions and $c(\cdot)$ is a function of the amount of time spent learning that the other person exists, traveling to their location, and talking to them.

Suppose that $E[f + g]$ is low mean and high variance for interactions with new (unknown) peers. Then our agent maximizes utility by soaking up lots of local, low cost social interactions. Once she knows someone well, which raises $E[f + g]$, then it pays to continue to interact with that friend even if the friend moves far away. This concept appears to describe our results as well as those of Festinger, Schachter, and Back [1963].

The alternative hypothesis (which we reject) is that our agent can predict with some certainty who would be a good future friend or partner for social interaction. If this were true, then she would probably be likely to travel across campus to meet someone new if that person was a good future prospect.

Suppose that interacting with a student of a different race is

5. In theory, student X could walk .7 miles to another part of campus to find out if some other dorm might house a previously unknown peer who can help him with his calculus problem set, or a new friend who wants to have dinner. But making this trip with no additional information, would be a costly and probably embarrassing thing for X to do, particularly if the probability of success is low.

more costly than interacting within race.[6] Since the expected benefit of interacting with any unknown person is small, even a small additional cost associated with cross-race interaction could have a large effect on the initiation of cross-race friendships. And this racial barrier would be self-perpetuating since in our model people derive utility from interacting with the same person repeatedly. Thus, even small costs associated with race could create a barrier to new social interactions that works in the same manner as geographic distance.

There is also a potential free rider problem; everyone in the society might agree that more interracial interaction would lower the costs of such interaction for everyone. But as an individual I may ignore this social benefit from my activities. This could explain why on modern university campuses students express both public and private support for reduced social segregation, and yet high levels of segregation persist (See Reid [2005]).

I.C. Empirical Framework

One goal of our analysis is to estimate the relative importance of geographic distance, racial similarity, family background, and common interests in determining who interacts with whom.[7] We do this by forming all possible pairs of students and asking who emails whom and with what intensity. We run Poisson regressions of the following form:

$$E\ (\#\text{ of emails between person 1 and person 2}) = e^{\mathbf{X}\beta},$$

where $\mathbf{X}\beta = \alpha + \mathbf{B1}$*(dummies for person 1's race, varsity athlete status, gender, Greek status, type of high school, financial aid status) + $\mathbf{B2}$*(dummies for person 2's race, varsity athlete status, gender, Greek status, type of high school, financial aid status) + $\beta 3$*(dummy for same graduating class) + $\beta 4$*(dummy for same freshman dorm) + $\mathbf{B5}$*(interactions of race and same freshman dorm) + $\mathbf{B6}$*(interactions of female with same dorm dummy, race dummies).

Here we combine into a single data point the volume from person A sending to person B and B sending to A. But we

6. Indeed the neuroscience literature suggests that White-Black interaction is more stressful than within-race interaction as measured in functional magnetic resonance imaging (fMRI) scans [Richeson et al. 2003].

7. We then take these estimates and ask questions, such as, a) how could policy-makers increase levels of interaction across diverse groups, and b) does a group means approach to peer effects adequately describe the social influences experienced by these students.

obtained similar results when we kept A to B and B to A as two distinct observations.

We run Poisson regressions for two related reasons. First, sending or receiving email is a rare, binomial event which can occur in any of the instants in time during the sample period. We are summing up over many instants in time (the sample period is thirteen months long). Thus, the number of emails should have a Poisson distribution. The histogram of the data supports this conjecture. And when we look at "effects" of right-hand-side variables, the effects appear to be multiplicative rather than additive suggesting a Poisson or other semi-logarithmic functional form. For example, putting two people in the same dorm and graduating class generally multiplies the expected number of emails by a factor of roughly 10, even for subgroups that have a very different baseline expected number of emails.[8]

In our regression tables we report the regression coefficients, standard errors, and $e^{\wedge}$(coefficient). The latter tells us how much the expected number of emails changes multiplicatively if the right-hand-side variable increases by one. (Since Poisson regression fits the number of emails to the form $e^{\mathbf{x}\beta}$, increasing x by one multiplies the predicted value by e^{β}.) For example, the coefficient on "same freshman dorm" is estimated to be 1.3, which means that being in the same freshman dorm raises the number of emails by a factor of 3.7.[9]

I.D. Do Emails Equal Friendship?

A natural question to ask is whether our email measure captures friendship or at least a meaningful level of social interaction. We believe strongly that within the campus we are studying the answer is yes. This conclusion is based upon our own experiences and a formal survey of friendship that we conducted in order to validate the emails measure.

Use of the Blitz (Dartmouth's email) system on campus is pervasive. Virtually all planned face-to-face interaction (for students, faculty, or staff) is organized over Blitz. Blitz is designed specifically to deliver intracampus email messages instanta-

8. We are greatly indebted to both referees for suggesting that we switch from OLS to Poisson.

9. Researchers often rely on the approximation that Poisson coefficients roughly correspond to percentage changes in the dependent variable. But in our case many coefficients of interest are too large in absolute value for this approximation to be helpful.

neously, so it serves the same purpose as Instant Messaging software which is popular in other organizations. In fact, pairs of roommates or faculty members with offices on the same hallway blitz (email) each other a great deal.

In order to demonstrate the strong positive correlation between friendship and email volume between two students, we surveyed a small subset of the subjects in our data set. We asked students to name their five closest friends on campus, knowing that we could then match this list against their emailing patterns. We emailed out 300 surveys and received back roughly 105 responses. Thus (for this validation exercise), we have a list of close friends for 105 students and email volumes between each of these 105 students and all other students.

If a student A considers student B to be a close friend, then there is a 75 percent chance that ten or more emails are sent between A and B during the thirteen-month period of the study. And the average volume between A and B is 136 messages. If A does not consider B to be a close friend, then there is a .2 percent chance that the total volume of emails exceeds ten, and the average volume is .21 messages. In short, we observe a strong connection between email volumes and self-reported friendship.

There are of course important caveats to this conclusion. First, our small survey of friends has only a 30 percent response rate and response is biased toward heavier users of Blitz. So the connection between emails and friendship may be less strong in the rest of the population. And smaller email volumes between two students might indicate a working relationship or brief exchange of information (e.g., times or dates for a meeting) rather than a friendship. Readers of this paper can substitute the term social interaction for friendship and hopefully still find the results meaningful.[10]

II. Data Description

We have the number of email messages sent and received among our users during June 2002 through July 2003. The data on email volumes are from Dartmouth's Netblitz email system. NetBlitz is the web-based version of Dartmouth's email software

10. We were deliberately bold in choosing the paper title because we believe that the qualitative results generalize beyond simply understanding email volumes among students at an elite college.

and is frequently used by students and alumni whenever they are off campus. To be included in the study as a primary user, a student must have used NetBlitz to check or send email at some point during the sample period, and a large fraction of students did so (see below). A single use of NetBlitz gives us access to that student's entire Dartmouth email history, whether the messages were created with NetBlitz, or Dartmouth's standard mail utility (Blitz), or any other email software. We recorded the number of email messages between any two students on the system between June 1, 2002, and July 31, 2003.

Whenever a student logged into NetBlitz, and had agreed to participate in the study, we captured ID numbers and volumes for the senders and recipients of their messages from the Inbox, Sent Messages folder, and any other folders that the student maintained. Thus, a single use of NetBlitz provides us with reams of data from the student's on and off campus email use.[11] A given message could be picked up from the sender's account, the recipient's account, or both. Our algorithm avoids double counting and distinguishes between sent and received messages.

We label the students using NetBlitz as primary users. Secondary users are those Dartmouth students who do not use NetBlitz but who appear in the data set by virtue of sending (or receiving) an email to (or from) a primary user. A sufficient but not necessary condition to capture all of a primary user's correspondence is that the user logs in to NetBlitz every six months. We drop the few primary users that logged in less frequently than every six months.

We dropped all emails that were sent to more than one person. Though such emails are often sent among friends, the emails also are sent to working groups and large organizations in which the individual members may have little interest or interpersonal interaction.

Numerous steps were taken to protect the human subjects in the study. First, as the researchers, our copy of the data did not include names but rather unique randomly assigned ID numbers. Second, no information on the content of the email messages was ever collected; we merely collected numbers of messages sent and received. Third, all subjects were given informed consent and the opportunity to opt out the study.[12]

11. In other words, we capture not just volumes sent during that particular session, but any information in the student's folders.

12. Eight percent of NetBlitz users opted out of the study.

TABLE I

FREQUENCY TABULATION OF PRIMARY USERS BY GRADUATING CLASS

Primary users are those who use the NetBlitz system for email and have agreed to participate. We have the full census of emails sent and received (not just on NetBlitz) for all primary users. Our data set includes nearly half of the classes of 2003 and 2004. The analysis that follows considers emails sent between primary users and all other students on campus. The set of primary users is large enough that all students appear at least once in the data set, by virtue of exchanging email with a primary user.

Sender's class	Frequency	Fraction of class
2003	491	0.47
2004	416	0.39
2005	236	0.23
2006	107	0.11
Total	1250	

In addition to data on email volumes, we also collected for each student the following data items: SAT scores, name of high school attended, financial aid status, race, gender, Dartmouth GPA as of July 2003, participation in varsity athletics, and a binary variable for membership in a fraternity, sorority, or coed Greek organization. These data are all available in Dartmouth's Banner database. We are indebted to the Computing Services Department for merging these student characteristics with our email data using student ID numbers.

Table I shows a tabulation of the primary users by graduating class. Nearly half of the graduating class of 2003 (491 out of 1050) and 39 percent of the class of 2004 are primary users. Essentially everyone else in these classes is a secondary user, because everyone communicated with one or more primary users during the thirteen-month period of the study. The percentage of primary users is smaller in the classes of 2005 and 2006 for two reasons. First, these classes spent less time off campus during the sample year and therefore had less need of NetBlitz. Second, Marmaros had previously made a point of advertising (via email) the availability of NetBlitz to the two older classes.

Table II shows some summary statistics on the primary users in the data set. On average, the group is 48 percent male, 72 percent White, and 49 percent had joined a fraternity or sorority by June of 2003. Table II also shows averages for several measures of academic ability including incoming math SAT score and

TABLE II
SUMMARY STATISTICS FOR PRIMARY USERS

Primary users are those who use the NetBlitz system for email and have agreed to participate. N = 1250. The analysis that follows considers emails sent between primary users and the set of all students.

Variable	Mean	Std. dev.	Min	Max
Male	0.48	0.50	0	1
Member of fraternity/sorority	0.49	0.50	0	1
White (0-1)	0.72	0.45	0	1
Black	0.05	0.21	0	1
Asian	0.14	0.35	0	1
Hispanic	0.06	0.24	0	1
Academic index (from admissions)	215.30	13.00	159.67	240.00
Cumulative GPA (as of 7/03)	3.38	0.35	1.78	3.99
Combined SAT Score	1427.43	103.89	1000	1600
Total messages sent and received	853.51	877.68	63	8737
Attended NYC specialized high school	0.03	0.17	0	1
Attended prep school with strong Dartmouth connection	0.05	0.22	0	1
Receives financial aid	0.53	0.50	0	1
Has 1 or more Black freshman roommates	0.08	0.27	0	1
Percent freshman floor Black	0.06	0.08	0	1
Percent freshman dorm Black	0.06	0.05	0	0.2
Number of freshman dormmates	34.49	17.63	0	91
Number of freshman hallmates	9.93	4.88	0	25

incoming Academic Index. The Academic Index can range from 60 to 240 and is a weighted average of SAT I scores (weight = ⅓), SAT II scores (weight = ⅓), and rescaled high school class rank (weight = ⅓). On average, the primary users exchanged (sent or received) 853 messages during the sample period with a standard deviation of 877 messages.

Three percent of primary users graduated from a New York City specialized ("exam") high school, meaning Stuyvesant, Brooklyn Tech, Bronx Science, or Hunter College High School.[13] Five percent of primary users attended one of the well-known, selective private schools that serves as a major feeder high school

13. Technically, Hunter College High School is not one of the traditional specialized high schools that use the same standardized test to admit students (Hunter has its own exam). In practice, Hunter is the one of the top, selective New York City public high schools that sends many students to Dartmouth.

for Dartmouth.[14] We use these two high school dummies as indicators to tell us something about a student's background. Fifty-three percent of primary users are financial aid recipients.

Each student has roughly 35 other freshman in their randomly assigned freshman dorm. On average, there are ten other freshman on one's freshman hallway.

A natural question to ask is whether the primary users are representative of Dartmouth students as a whole. We address this question in Appendix 1 where we compare the primary users with everybody else (i.e., the secondary users). The primary users are much more likely to be members of a fraternity or sorority, but this discrepancy is partly due the fact that only 107 freshman are primary users and the freshman are prohibited from joining. When we compare primary and secondary users who are juniors or seniors, the difference in fraternity membership is no longer statistically significant.

Average math SATs for the primary users and secondary users are similar at 715 and 706, respectively. And cumulative GPA is similar across the two groups. Overall, we believe that our group of primary users is large enough and diverse enough to enable us to form conclusions about the behavior of Dartmouth students as a group, even though primary users select into our sample by choosing to use NetBlitz.

Black students are significantly underrepresented among the primary users; 4.9 percent of primary users are Black versus 8.0 percent of the secondary users. If the Black secondary users have significantly different patterns of interaction than the Black primary users, our results may not generalize across the two groups. Even given this bias, we are still well positioned to study cross-race interaction because the only emails that do not enter our data set are those exchanged between two secondary users. Suppose that Black students never used Netblitz but that all White students did. We would end up with the complete census of White-Black email interactions since the set of secondary users includes everyone.

We examine the volume of emails exchanged between each

14. We created a dummy equal to 1 for students from private prep schools that fall within the top twenty feeders to schools to Dartmouth (in recent years). The list of schools includes Andover, Exeter, Walt Whitman (MD), Lawrenceville, St. Paul's, Deerfield, St. Ann's (NY), Horace Mann, Punahou, Winsor, Trinity (NY), Buckingham, Browne, and Nichols, Dalton School, Pingry School, Loomis Chafee, and Collegiate.

TABLE III

Pair Level Summary Statistics

We consider all possible pairings of primary users and all Dartmouth students. The analysis that follows looks at the volume sent (if any) between each primary user and every student on campus. In the labels below, person 1 is the primary user, and person 2 is the other person in the pair (be they a primary or secondary user).

Variable	Obs.	Mean	Std. dev.	Min	Max
Talks (0–1) (i.e., email each other at least once)	4,225,623	0.007	0.083	0	1
Email each other at least 5 times each way	4,225,623	0.002	0.048	0	1
Volume sent between persons 1 and 2	4,225,623	0.253	9.831	0	3745
Volume sent conditional on volume > =1	29,197	32.381	112.910	2	3745
Persons 1 and 2 are members of same class year	4,225,623	0.237	0.425	0	1
Same freshman floor	4,225,623	0.003	0.054	0	1
Same freshman dorm building	4,225,623	0.010	0.101	0	1
Same freshman year cluster of buildings	4,225,623	0.031	0.172	0	1
Distance between freshman rooms in thousands of feet	4,225,623	1.421	0.839	0.060	3.273
Same major	4,225,623	0.069	0.254	0	1

primary user and all other students (summing over emails going in either direction). To do this, we form all possible pairwise combinations of a primary user and any student in the data set (be they a primary or secondary user). We cross the set of 1250 primary users with 4000+ students which results in 5.3 million pairs of students. We then eliminate duplicate observations resulting from A to B and B to A potentially being counted as separate pairs leaving us with 4.2 million unique pairs. Any of these 4.2 million possible social connections might be active (i.e., have email traffic), though in fact about 29,000 of the connections are active during the sample period.

For the tables and associated discussion, we label the first person in the pair (who is always a primary user) as "Person 1" and the second person in the pair (who may be a primary or secondary user) as "Person 2."

Table III shows summary statistics at the pair level. In the first row, we see that in .7 percent of the pairs, person 1 has sent one or more emails to person 2 *and* received one or more emails from person 2.[15] In .2 percent of pairs five or more emails have traveled in each direction (for a total volume of ten or more).[16] Conditional on having sent and received email from the other person in the pair, 32.4 messages are sent on average, though the standard deviation is 113 messages and the maximum volume in a pair is over 3700 messages. Roughly 24 percent of the pairs consist of two members from the same graduating class. Of the pairs .3 percent are from the same class and the same freshman hallway. One percent of pairs are from the same class and same freshman dorm.

The pair members' relative location freshman year is important for several reasons. First, we show that there is an incredibly strong correlation between freshman year housing assignment and the likelihood (and intensity) of person 1 emailing person 2. This connection remains strong even after graduation. Second, because freshman dorms and hallways are randomly assigned (as in Sacerdote [2001]), we can give this correlation a causal interpretation.[17]

III. Results

In Table IV we examine how the amount of social interaction between two students varies by the race of the students and whether or not they are in the same entering freshman dorm. We limit the sample to pairs from the same entering class and for which person 1 is White. The top row shows that White-non-Black pairs not in the same freshman dorm exchange .71 emails on average.[18] If both people are in the same dorm, the mean jumps to 2.95, indicating a multiplicative effect from being in the same dorm of 4.2×.

15. We define an indicator variable called "talks" which equals 1 if persons 1 and 2 have sent and received emails from each other.

16. Our empirical results are similar whether our dependent variable is the actual volume, or a dummy for two or more emails sent (total) or a dummy for ten emails sent (total).

17. The housing office takes all of the freshman housing applications and separates them into several groups based on gender and self-reported smoking, neatness, and sleeping habits. The groups are then shuffled. Groups of roommates are created randomly within a pile. Floormates and dormmates are drawn randomly across piles.

18. In results not reported, we limited the sample to White-Black plus White-White pairs and found similar means.

TABLE IV

MEAN VOLUMES AND PROBABILITIES OF TALKING BY RACE AND GEOGRAPHY

We limit the sample to pairs from the same class and in which person 1 (the primary user) is White. We stratify by the second person being Black and by being in the same freshman dorm. For each pair we compute the mean volume (number of messages sent back and forth) and the probability of talking (i.e., sending at least one message in each direction).

Person 2 Black?		Same dorm? No	Same dorm? Yes	Ratio (same dorm to not)
No	Mean (Volume)	0.708	2.947	4.162
	Mean (Talks?)	0.017	0.054	3.176
	N	643,294	28,874	
Yes	Mean (Volume)	0.212	0.619	2.920
	Mean (Talks?)	0.008	0.021	2.625
	N	48,182	2,220	

When person 1 is White and person 2 is Black, the multiplicative effect of being in the same freshman dorm is 2.92, but measured from a much lower base level of interaction; the mean volume between a randomly chosen Black student and a randomly chosen White classmate is .21 messages for students in different dorms and .62 for students in the same dorm.

Next we consider the race effect by comparing rows 1 and 4. Holding constant that both students are in the same dorm (column (2)) and that person 1 is White, the second person being Black reduces mean email volume by a factor of 4.76 (i.e., 2.95/.62). For students not in the same dorm, the second student being Black reduces email volume by a factor of 3.34.

Table IV also shows the mean probability that the two people exchange at least one email in each direction ("talk"). Moving a White-Black pair to the same dorm increases the probability of talking by a factor of 2.63.

We draw several conclusions from this analysis. First, the effects of race, geography, and being from the same cohort are large. White-non-Black pairs have roughly 3–5 times more interaction than White-Black pairs. When two people are from the same graduating class, being in the same dorm raises the amount of interaction by a factor of 3 or 4. For White-Black pairs, the positive effect of being in the same dorm is nearly large enough to offset the negative race effect; White-Black pairs in the same dorm have mean email volumes of .62 versus .71 for White-non-Black pairs not in the same dorm.

Second, the analysis of means suggests we should model the effects from student characteristics using a multiplicative form (e.g., a Poisson regression) rather than a linear form. The same dorm effect is on the order of a 3–4× increase whether we start from the baseline level of a mixed race pair or the higher base level of a same race pair. The race effect is on the order of a 3× effect whether we consider interaction within versus across freshman dorms.

We next ask how much having a Black freshman roommate or a greater than median percentage of Black students on one's freshman hallway increases a student's interaction with Black students. For all the non-Black students, having a Black roommate raises the fraction of emails volume exchanged with Black students from 4.4 to 4.9 percent.[19] This only amounts to an extra three emails exchanged with Black students. We show later in the paper that this increase in emails can easily be accounted for by the extra email volume between White students and their own Black roommates. The effects are not suggestive of a multiplier effect in which Black roommates introduce White students to new networks of Black students. Living on a hallway with more than median percent Black raises a non-Black student's fraction of emails exchanged with Black students from 4.2 percent to 4.7 percent.

In Table V we limit the sample to White students and run regressions at the student level. We regress the student's percent of email volume exchanged with Black students on the percent Black in her dorm, her hallway, and a dummy for having any Black roommate.[20] We control for all observed student characteristics including gender, SAT scores, financial aid status, type of high school attended, Greek, and athletic status. Using column (1), a 10 percent increase in the percent Black in a student's dorm leads to a .1 percent increase in volume exchanged with Black students, which translates to .9 messages. This effect is fully explained by the same dorm effect for any two students, and need not imply that having Black dormmates increases a White student's likelihood of interacting with Black students outside her

19. We set the dummy for having a Black roommate equal to one if one or *more* roommates are Black.

20. The dorm and hallway percent Black calculation excludes the student herself.

TABLE V

STUDENT LEVEL REGRESSIONS OF FRACTION OF INTERACTIONS THAT ARE CROSS-RACE ON RACIAL COMPOSITION OF DORM, HALLWAY, ROOM

We limit the sample to White students. For each student we compute the fraction of their interactions that take place with Black students. We regress this fraction on the percent Black on the student's freshman hallway, percent Black in freshman dorm, and a dummy for having a Black roommate. We also control for the student's characteristics including race, SAT scores, gender, fraternity membership, financial aid status, athlete status, and type of high school attended (coefficients not reported).

	(1) Fraction volume with Black students (Person 1 White)	(2) Fraction volume with Black students (Person 1 White)	(3) Fraction volume with Black students (Person 1 White)
Percent dorm Black	0.095		
	(0.040)*		
Percent hallway Black		0.014	
		(0.024)	
Black roommate			0.002
			(0.007)
Constant	0.082	0.083	0.083
	(0.032)*	(0.032)*	(0.032)*
Observations	900	900	900
R^2	0.021	0.015	0.015

Standard errors are in parentheses.
* Significant at 5 percent; ** significant at 1 percent.

dorm.[21] In column (3) the point estimate indicates that having a Black freshman roommate raises a White student's percent of emails sent to Blacks by .2 percent and the effect is not statistically significant.

In Table VI we arrange the data at the student-pair level and run Poisson regressions of total emails exchanged on the characteristics of both students in the pair. Each column is a separate regression. Each row shows the coefficient, the standard error in parentheses, and $e^{\wedge}$coefficient in square brackets [ ]. The purpose of the latter is to show the multiplicative effect of a one-unit

21. A 10 percent increase in the fraction Black would correspond to adding about three more Black students to the White student's dorm. Using the means for same dorm versus not in Table IV, we would expect an additional $.4 * 3 = 1.2$ messages.

TABLE VI

REGRESSION OF NUMBER OF EMAILS EXCHANGED (VOLUME) AND "TALKS" ON SENDER AND RECIPIENT CHARACTERISTICS

The dependent variables are a) the total number of emails sent or received between persons 1 and 2 and b) a dummy ("talks") which equals 1 if person 1 sent and received back one or more emails from person 2. The mean of the dummy variable is .007, meaning that there is .7 percent chance that a randomly chosen pair interacts. The mean number of emails exchanged is .253. Each column conditions on the race of person 1. White is always the excluded category for race. The pairs are structured so that person 1 is a primary user and person 2 is any other student.

Columns (1)–(5) are Poisson regressions. The number in square brackets [ ] is $e^{\wedge}$coefficient which is the multiplicative effect from a unit change in the right-hand-side variable. Column (6) is a probit, and partial derivatives are shown. Standard errors use clustering at the person 1 level. The purpose of column (1) is to show that the race and same dorm effects estimated via Poisson closely match the multiplicative effects calculated using the means of the raw data (shown in Table IV).

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of emails exchanged Person 1 is White (Poisson)	Number of emails exchanged Person 1 is White (Poisson)	Number of emails exchanged Person 1 is Black (Poisson)	Number of emails exchanged Person 1 is Asian (Poisson)	Number of emails exchanged Person 1 is Hispanic (Poisson)	Talks? (1+ each way) Person 1 is White (Probit) $\partial y/\partial x$
Person 2 same freshman dorm and class	1.419 (0.101)** [4.133]	1.295 (0.122)** [3.651]	1.366 (0.293)** [3.920]	1.573 (0.192)** [4.821]	1.277 (0.331)** [3.586]	0.013 (0.001)**
Black Person 2* same freshman dorm		−0.247 (0.293) [0.781]	−0.641 (0.449) [0.527]	0.886 (0.448)* [2.425]	−0.373 (0.477) [0.689]	−0.002 (0.001)**
Asian Person 2* same freshman dorm		0.521 (0.255)* [1.684]	0.095 (0.612) [1.100]	−1.228 (0.314)** [0.293]	−0.266 (0.526) [0.766]	0.001 (0.001)*
Hispanic Person 2* same freshman dorm		0.858 (0.350)* [2.358]	0.317 (0.618) [1.373]	−0.420 (0.450) [0.657]	−1.060 (0.597) [0.346]	0.001 (0.001)
Other non-White 2* same freshman dorm		−0.039 (0.383) [0.962]	0.488 (0.379) [1.629]	0.534 (0.448) [1.706]	−0.332 (0.734) [0.717]	0.002 (0.001)
Person 2 is Black	−1.209 (0.100)** [0.298]	−1.019 (0.144)** [0.361]	2.816 (0.441)** [16.710]	−0.337 (0.365) [0.714]	−0.166 (0.289) [0.847]	−0.002 (0.000)**
Person 2 is Asian		−0.569 (0.116)** [0.566]	−0.058 (0.285) [0.944]	1.539 (0.160)** [4.660]	−0.435 (0.193)* [0.647]	−0.002 (0.000)**

TABLE VI
(CONTINUED)

	(1) Number of emails exchanged Person 1 is White (Poisson)	(2) Number of emails exchanged Person 1 is White (Poisson)	(3) Number of emails exchanged Person 1 is Black (Poisson)	(4) Number of emails exchanged Person 1 is Asian (Poisson)	(5) Number of emails exchanged Person 1 is Hispanic (Poisson)	(6) Talks? (1+ each way) Person 1 is White (Probit) $\partial y/\partial x$
Person 2 is Hispanic		−0.217 (0.097)* [0.805]	0.683 (0.234)** [1.980]	0.317 (0.237) [1.373]	0.992 (0.205)** [2.697]	0.000 (0.000)**
Person 2 is other non-White		−0.533 (0.179)** [0.587]	0.111 (0.475) [1.117]	−0.802 (0.210)** [0.448]	−0.003 (0.468) [0.997]	−0.002 (0.000)**
Same class year	1.898 (0.064)** [6.673]	1.762 (0.068)** [5.824]	1.575 (0.149)** [4.831]	1.965 (0.132)** [7.135]	1.608 (0.156)** [4.993]	0.009 (0.000)**
Varsity athlete (Person 1)		−0.613 (0.096)** [0.542]	−0.350 (0.377) [0.705]	−0.405 (0.221) [0.667]	−0.962 (0.245)** [0.382]	−0.002 (0.000)**
Varsity athlete (Person 2)		−0.566 (0.105)** [0.568]	−0.731 (0.207)** [0.481]	−1.071 (0.129)** [0.343]	−0.547 (0.162)** [0.579]	−0.002 (0.000)**
Both are athletes		1.297 (0.142)** [3.658]	1.296 (0.576)* [3.655]	1.553 (0.343)** [4.726]	1.019 (0.355)** [2.770]	0.010 (0.001)**
Greek member (Person 1)		−0.682 (0.097)** [0.506]	−0.042 (0.218) [0.959]	−0.179 (0.217) [0.836]	−0.319 (0.247) [0.727]	−0.003 (0.000)**
Greek member (Person 2)		−0.207 (0.097)* [0.813]	−0.514 (0.190)** [0.598]	−0.308 (0.163) [0.735]	−0.107 (0.250) [0.899]	−0.001 (0.000)**
Both are in Greek organizations		1.382 (0.131)** [3.983]	1.859 (0.279)** [6.417]	1.008 (0.254)** [2.740]	0.842 (0.322)** [2.321]	0.010 (0.001)**
Same major		0.239 (0.091)** [1.270]	0.093 (0.166) [1.097]	0.097 (0.194) [1.102]	0.175 (0.202) [1.191]	0.002 (0.000)**
Absolute difference in SAT scores		−0.003 (0.000)** [0.997]	−0.001 (0.001) [0.999]	−0.001 (0.001) [0.999]	0.000 (0.001) [1.000]	0.000 (0.000)**
Person 1 is male		−0.150 (0.069)* [0.861]	0.098 (0.209) [1.103]	−0.148 (0.138) [0.862]	0.166 (0.189) [1.181]	0.000 (0.000)
Person 2 is male		−0.333 (0.049)** [0.717]	−0.375 (0.151)* [0.687]	−0.393 (0.109)** [0.675]	−0.572 (0.134)** [0.564]	−0.001 (0.000)**

TABLE VI
(CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of emails exchanged Person 1 is White (Poisson)	Number of emails exchanged Person 1 is White (Poisson)	Number of emails exchanged Person 1 is Black (Poisson)	Number of emails exchanged Person 1 is Asian (Poisson)	Number of emails exchanged Person 1 is Hispanic (Poisson)	Talks? (1+ each way) Person 1 is White (Probit) $\partial y/\partial x$
Both are male		−0.122 (0.050)* [0.885]	0.086 (0.154) [1.090]	0.030 (0.109) [1.030]	0.005 (0.137) [1.005]	0.003 (0.000)**
Person 1 went to NYC exam school		0.565 (0.270)* [1.759]	−0.131 (0.178) [0.877]	−0.647 (0.250)** [0.524]	−1.259 (0.289)** [0.284]	0.002 (0.001)*
Person 2 went to NYC exam school		−0.052 (0.140) [0.949]	−0.045 (0.333) [0.956]	−0.071 (0.301) [0.931]	0.341 (0.323) [1.406]	0.000 (0.000)
Both 1 and 2 went to NYC exam schools		1.779 (0.643)** [5.924]	0.389 (0.522) [1.476]	1.369 (0.491)** [3.931]	0.678 (0.359) [1.970]	0.003 (0.002)
Person 1 went to fancy prep school		−0.247 (0.116)* [0.781]	−0.045 (0.216) [0.956]	−0.725 (0.239)** [0.484]	0.026 (0.255) [1.026]	0.000 (0.000)
Person 2 went to fancy prep school		0.173 (0.178) [1.189]	−0.563 (0.222)* [0.569]	0.079 (0.190) [1.082]	−0.008 (0.246) [0.992]	0.000 (0.000)
Both 1 and 2 went to fancy prep school		0.536 (0.284) [1.709]	−0.320 (0.734) [0.726]	−0.751 (0.731) [0.472]	0.086 (0.651) [1.090]	0.004 (0.001)**
Person 1 on financial aid		−0.175 (0.093) [0.839]	−1.101 (0.346)** [0.333]	0.148 (0.213) [1.160]	−0.230 (0.264) [0.795]	−0.001 (0.000)**
Person 2 on financial aid		−0.294 (0.075)** [0.745]	−0.583 (0.408) [0.558]	0.241 (0.221) [1.273]	−0.848 (0.260)** [0.428]	−0.001 (0.000)**
Both on financial aid		0.393 (0.113)** [1.481]	1.074 (0.469)* [2.927]	0.028 (0.261) [1.028]	1.032 (0.300)** [2.807]	0.001 (0.000)**
Person 2 is Black* Person 1 has Black Roommate		−0.291 (0.312) [0.748]				−0.001 (0.001)
Person 2 is Black* % Black in 1's freshman dorm		1.516 (1.749) [4.554]	−3.501 (3.765) [0.030]	1.624 (4.063) [5.073]	4.791 (2.681) [120.422]	0.002 (0.004)

TABLE VI
(CONTINUED)

	(1) Number of emails exchanged Person 1 is White (Poisson)	(2) Number of emails exchanged Person 1 is White (Poisson)	(3) Number of emails exchanged Person 1 is Black (Poisson)	(4) Number of emails exchanged Person 1 is Asian (Poisson)	(5) Number of emails exchanged Person 1 is Hispanic (Poisson)	(6) Talks? (1+ each way) Person 1 is White (Probit) $\partial y/\partial x$
Person 1's percent dorm Black		−1.128 (0.752) [0.324]				−0.003 (0.002)
Person 1 has a Black roommate		0.181 (0.144) [1.198]				0.000 (0.000)
Constant	−2.242 (0.054)**	−1.140 (0.103)**	−2.129 (0.437)**	−2.382 (0.258)**	−1.666 (0.347)**	
Observations	3,021,484	2,923,120	192,279	560,120	245,165	2,923,120

Robust standard errors are in parentheses. * significant at 5 percent; ** significant at 1 percent. NYC exam schools included are Stuyvesant, Bronx Science, Brooklyn Tech, and Hunter College High School. Fancy prep schools include Andover, Exeter, Walt Whitman (MD), Lawrenceville, St. Paul's, Deerfield, St. Ann's (NY), Horace Mann, Punahou, Winsor, Trinity (NY), Buckingham, Browne and Nichols, Dalton School, Pingry School, Loomis, Chafee, and Collegiate. These are the private high schools that are within the top twenty feeder high schools to Dartmouth.

change in the right-hand-side variable. The omitted categories for person 1 and 2's race is always White. The interaction dummies are constructed such that the coefficients should be multiplied together to calculate the total effect. For example, the interaction effect for person 2 being "Black and in the same class and dorm" is on top of the baseline effects of "same freshman dorm" and "same class."

In column (1) we limit the sample to pairs in which person 1 (the primary user) is White. We regress the number of emails on dummy variables for person 2 being Black, being in the same freshman dorm, and being in the same freshman class. We show this simplified specification mainly to verify that the Poisson regressions reproduce the same effects that we observed in the means of the raw data in Table IV. The regression shows that being in the same freshman dorm multiplies the number of emails by 4.13. Table IV indicates a multiplicative effect of 4.16. If person 2 is Black, the expected number of emails is multiplied by .299 (a 70 percent reduction), which is also consistent with Table IV.

Column (2) again limits the sample to pairs in which person 1 is White and includes all characteristics for both students in the pair. Many of our key results can be seen in this column. Being in the same class year has a multiplicative effect of 5.8.

We include all interactions of person 2's race and same freshman dorm.[22] The omitted category for person 2's race is always White. For White-White pairs, being in the same freshman dorm has a multiplicative effect of 3.7 (on top of the same class effect). Interestingly, the interactions of same dorm with race are generally not significantly different from the White-White same dorm effect. In other words, geographic proximity enhances cross-race interactions to the same degree that proximity enhances within race interactions.

However, there are large negative level effects from person 2 being Black when person 1 is White. Overall, White-Black pairs are estimated to have 64 percent fewer interactions (1 minus the .36 shown in square brackets for the "person 2 is Black" estimate). White-Asian pairs have 44 percent fewer interactions, and White-Hispanic pairs have 20 percent fewer interactions.

We turn now to the effects of both students in the pair having similar interests including athletic participation, fraternity and sorority membership, and college majors. Again considering pairs where person 1 is White, being in the same college major raises the number of interactions by 27 percent. This effect is not statistically significant when person 1 is Black, Asian, or Hispanic (columns (3)–(5)).

Being an athlete has a negative level effect on interaction, whether we examine the coefficient for person 1 being an athlete or person 2 being an athlete. However, the interaction term for both students being athletes has a multiplicative effect of 3.7 on the number of emails. If person 1 is an athlete and person 2 is not, we find 46 percent fewer emails relative to the base category of nonathlete-nonathlete pairs. However, athlete-athlete pairs enjoy 13 percent more emails than the base category.[23] The effects for Greek membership work in much the same way. If only person 1 is Greek, there are 49 percent fewer emails relative to non-Greek-non-Greek pairs. If both 1 and 2 are Greek, we see 64 percent more emails relative to the base category. Since athletic

22. "Same freshman dorm" always implies same class.

23. Here we multiplied the three relevant effects from person 1 is an athlete, person 2 is an athlete, and the interaction term to get the total effect.

status, Greek status, and majors are all determined endogenously, we offer these as descriptive statistics rather than causal effects. Furthermore, we do not know whether athletes are talking mostly to individuals on their same team or to athletes on other teams. And Greek members may simply be interacting a great deal with members of their own organization and not with members of other organizations.

In the next rows we ask whether differences in academic ability reduce the amount of social interaction. We take combined SAT score (math plus verbal) as a measure of pretreatment academic ability. We look at the effect of the absolute difference in SAT score between person 1 and person 2, controlling for the SAT score for both students. Differences in SAT scores do reduce the amount of interaction between the pair, but the effect is modest in size. A 200 point difference in SATs results in a 0.6 percent reduction in email volume. This effect is only statistically significant when person 1 is White. The effect is statistically insignificant when person 1 is Black, Asian, or Hispanic. We tried many alternative specifications such as creating dummies for person 1 and person 2's quartiles of SAT and running the fully interacted specification. We found no evidence that a particular combination of SAT scores stimulated or hindered social interaction among students.

We also examine the effects of family background using dummy variables for three different characteristics. We know whether each student is on financial aid, whether they attended one of New York's specialized (exam) high schools (Stuyvesant, Bronx Science, Brooklyn Tech, Hunter College High), and whether they attended one of the elite private high schools that is among the major feeder schools to Dartmouth. Student pairs who both went to a NY exam school exchange 9× as many emails as student pairs in the base category. The effect of both students attending an elite private school is smaller and not statistically significant.

Some of the large effect on both students attending NY specialized high schools probably stems from the students in the pair knowing each other before enrolling at Dartmouth. If true, this would be consistent with our model of friendship formation in which once a social connection is established, people derive utility from interacting with the same person over and over again.

If one of the students in the pair is a financial aid recipient and not the other, the number of emails is reduced by 20 percent, relative to the base category of nonaid–nonaid. If both students receive financial aid, then the amount of interaction is roughly equal to the base category. These results are from column (2) which uses pairs where person 1 is White. When person 1 is Black, Asian, or Hispanic (columns (3)–(5)), no clear pattern emerges from the effects of financial aid status on the level of interaction.

The final rows of Table VI address whether having a Black dormmate or roommate affects my interracial interactions with Black students outside my dorm. For pairs in which person 1 is White, we include dummy for person 1 having a Black freshman roommate and the percent Black in 1's freshman dorm. We then interact these characteristics for person 1 with a dummy for whether person 2 is Black. Having a Black roommate or a high percentage Black in one's dorm has no statistically significant effect on a White student's volume of interactions with Black students in general.[24] This is consistent with the student level analysis we showed in Table V.

A separate question is whether or not minority students become socially isolated if they are grouped together in the same freshman rooms or dorms. In results not reported here we find that for Black students, having a Black roommate or increased percent Black in one's dorm does not increase the volume of interactions with *other* Black students outside of one's dorm.

Columns (3)–(5) limit the sample to pairs in which person 1 is Black, Asian, and Hispanic, respectively. In general, many of the key results from column (2) remain. For example, there is a large same freshman dorm effect that increases the number of emails sent by a factor of between 3.5 and 5. The same dorm effect is largest for pairs in which person 1 is Asian. There is a large and positive same race effect for Black students. In column (3) where person 1 is Black, the coefficient on person 2 being Black is 2.816 which means that email volume is increased by a factor of 16.7. This same race effect is much larger than the same race attraction experienced by other groups.

In column (6) we switch the dependent variable to a dummy variable for person 1 having sent and received at least one email from person 2. The mean of this dummy is .007 meaning that .7 percent of pairs are in active email communication ("talking"). We limit the sample to pairs in which person 1 is White. We run a

24. There is still of course the direct effect that being close to any student of any race makes one more likely to interact with that particular student.

probit regression and report partial derivatives. We find results that are qualitatively similar to the Poisson regression in column (2). Being in the same freshman dorm raises the probability of talking by 1.3 percent which indicates that the same dorm effect increases the probability of talking by a factor of 2.9. The interactions between "same dorm" and person 2's race are generally small indicating that the same dorm effect on talking is the same magnitude both within and across races. White-Asian pairs have a slightly larger same dorm effect than White-White pairs while White-Black pairs have a slightly smaller effect. There are still large and statistically significant level effects from person 2's race. For White-Black pairs, the likelihood of talking falls by .2 percent relative to White-White pairs.

In Table VII we investigate more closely the effect of distance and how this effect changes from freshman to senior year. We limit the sample to pairs in which person 1 is White. To measure geographic distance in more detail, we include dummies for same freshman year room, floor (hallway), dorm, and cluster of dorms. At Dartmouth, a cluster of dorms is a collection of 2–4 buildings that are connected by common rooms, porches, or outdoor breezeways. We also include the physical distance between the freshman year rooms of person 1 and person 2, measured in thousands of feet. We include all the same student characteristics used in Table VI (race, gender, graduating class, fraternity membership, etc.) but only report coefficients on the distance measures.

Sharing the same freshman year dorm (but not the same hallway) doubles the number of emails sent (column (1)) and increases the probability of talking by .3 percent (column (2)). Sharing the same floor delivers a 2.3× effect over and above the 2.0× effect from being in the same dorm. Being in the same room adds an additional effect of 3.1×. This means that freshman year roommates share 14.3 times as many emails relative to two randomly chosen classmates who are not from the same freshman dorm.[25]

Being in the same cluster has no additional effect, nor does the linear measure of distance. This means that proximity only has a significant effect at very close distances. We show below that this fact limits the degree to which proximity can be used by policy-makers to create shifts in the overall amount intergroup

25. The regression controls for but does not report the effect from the two students being in the same class.

TABLE VII

THE EFFECT OF DISTANCE ON THE AMOUNT OF INTERACTION AND HOW THE EFFECTS OF DISTANCE DEGRADE WITH TIME

Sample is limited to pairs in which person 1 is White. The dependent variables are the total volume exchanged between person 1 and person 2 during September 2002–July 2003 and a dummy for "has sent and received 1 or more emails." Columns (1), (3), and (4) are Poisson regressions. Standard errors use clustering at the person 1 level. The number in square brackets [ ] is $e^{\wedge}$coefficient. This is the multiplicative effect on the predicted number of emails for a one unit change in the right-hand-side variable. Column (2) is a probit, and partial derivatives are shown. Column (3) is for freshmen, and (4) is for seniors. The point here is to see the degree to which the effects of freshman year distance degrade over time.

Regressions also control for (but suppress coefficients for) all *X*s in the previous table include race, financial aid, athletic, fraternity status, same graduating class, etc.

	(1) Total volume (Person 1 is White)	(2) Sent at least 1 email (Person 1 is White)	(3) Total volume (Person 1 is a freshman and White)	(4) Total volume (Person 1 is a senior and White)
Same room	1.134	0.034	1.857	0.971
	(0.160)**	(0.005)**	(0.312)**	(0.286)**
	[3.108]		[6.404]	[2.641]
Same freshman year floor	0.838	0.011	1.649	0.379
	(0.217)**	(0.001)**	(0.375)**	(0.375)
	[2.312]		[5.202]	[1.461]
Same freshman year dorm	0.661	0.003	0.289	0.925
	(0.225)**	(0.001)**	(0.336)	(0.408)*
	[1.937]		[1.335]	[2.522]
Same freshman year cluster of dorms	0.237	0.001	0.308	0.471
	(0.165)	(0.000)**	(0.403)	(0.316)
	[1.267]		[1.361]	[1.602]
Freshman year residential distance in thousands of feet	−0.088	0.000	−0.070	0.037
	(0.049)	0.000	(0.107)	(0.072)
	[0.916]		[0.932]	[1.038]
Observations	2,923,120	2,923,120	230,920	1,094,550

Robust standard errors are in parentheses.
* Significant at 5 percent; ** significant at 1 percent.

interaction that occurs. This result is confirmed in Figure I in which we graph the average volume of email exchanged among pairs against their residential distances freshman year. Distance is quite important but only at close distances.

In columns (3) and (4) we run the Poisson regression separately for freshmen and seniors. As one would expect, the fresh-

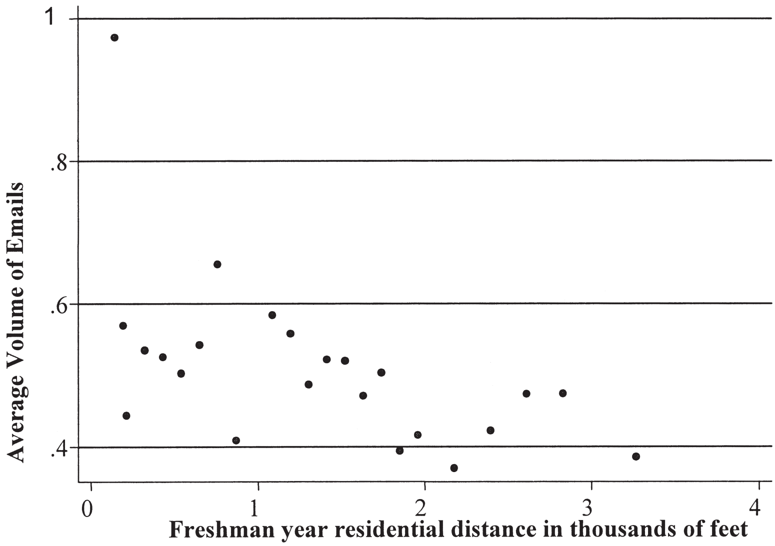

FIGURE I
Average Volume of Email Sent at Each Distance
The data consist of all pairs of students who entered Dartmouth in the same year. Distance refers to the distance between the two students' freshman rooms. Physical distance is grouped into 22 categories, and we take the average volume (including zeroes) across all pairs within a distance category.

men show larger effects from freshman year housing distance. In their first year, freshman roommates are exchanging 44.6 times more emails than two freshmen not in the same dorm. But interestingly, the effects of freshman housing are still strong three years later. As seniors, former freshman roommates are exchanging 9.8 times more emails than two randomly chosen seniors.

This is consistent with our proposed model of friendship in which chance meetings lead to long-term bonds between two people. At the time of enrollment, a student's list of Dartmouth social connections is a mostly blank slate. Dartmouth randomly creates a social network by housing freshman together, and this network shows a great deal of persistence. In results not shown here, we find that following graduation, former freshmen roommates and dormmates continue to be much more likely to email each other than randomly chosen classmates.

Furthermore, given the panel nature of the data, we are able to observe the volume of interaction between a pair of students during months when they are both on campus and months where one of the students is off campus. We find that email volumes are significantly higher when both students are on campus and hence able to interact face to face as well as through email. Our preferred interpretation of this fact is that email is a complement to face-to-face interaction. Students can use email to set up face-to-face interaction, or perhaps they get more utility from exchanging emails with someone with whom they have just seen or will see shortly.

III.A. *How Isolated Are Various Groups and Can Housing Policies Change This?*

Bowen and Levin [2003] raise the concern that within the College and Beyond Schools athletes may be an isolated group of students who fail to interact with the rest of the campus. As mentioned above, other authors and policy-makers have similar concerns about the degree to which minority students form isolated social groups. To investigate this, we calculate the fraction of correspondents and fraction of emails that are within group versus out of group for three segments of the population: football players, all athletes, and Black students. The student body is roughly 7 percent Black, 26 percent athletes, and 2.3 percent football players.

Black primary users share about 44 percent of their email volume with other Black students, whereas non-Black students share about 4 percent of their email volume with Black students. Athletes exchange 52 percent of their emails with other athletes. Most striking is the fact that football players exchange 30 percent of their emails with other football players despite being roughly 2 percent of the population.

Whether these numbers are high or low depends on the reader's priors. We can reject extreme hypotheses of isolation since Black students have more than half of their interactions with non-Black students. And athletes have nearly half of their interactions out of group.

One interesting question from a peer effects and policy perspective is the degree to which out of group interaction could be increased by creating further mixing within freshman dorms. Our results are not particularly encouraging if the goal is to generate further cross-race interaction.

We have tried various simulations of redistributing Black freshman across dorms to increase predicted Black-White interaction. We take the coefficients from the Poisson regression from

Table VI column (2). We reassign Black freshmen across dorms to deliver a uniform percent Black in each dorm, for a given class year. This is a meaningful exercise because the actual random assignment to groups of 30 to 50 yielded some dorms with as few as zero or as many as six Black students rather than the roughly three of a perfectly even distribution. Across the entire campus after reassigning the Black students' housing, we predict an additional 146 emails between White-Black pairs as compared with the current White-Black total volume of 16,800 messages. This indicates that each White student would exchange an additional .16 emails with Black students, or a less than 1 percent increase on the mean of 18–19 emails currently exchanged during the sample period.

The intuition for this finding is as follows: we found in Tables IV and VI that placing two freshmen in the same dorm increases their email volume by a factor of 3 or 4. However, in the case of White-Black interactions, that increase is from a low base of .2 emails. If we could give each White freshman an additional Black dormmate (say from an all Black dorm if one existed), she would experience the increase of .4 emails to that Black student shown in Table IV. But since we are redistributing the Black students, rather than creating a net increase, we get much less than the .4 effect.

An opposite policy experiment to consider would be complete segregation of the Black freshmen. For the average White student, this would be a loss of about 3 Black dormmates and would result in a reduction of $3 * .4 = 1.2$ emails exchanged with Black students. This is again modest relative to the current average of nineteen emails. Proximity has a strong positive effect, but only 3 percent of pairs in the sample share the same freshman dorm and class. On average, only about 9 percent or so of Black-White emails take place within one's freshman dorm group.

In contrast, changing the entire class's composition would have a large effect on my social interactions. A student interacts 5–7 times more with a student from her own class than with a student from a different class. When we add another minority student or high SAT student into a class, every student of 1050 in the class experiences a large jump in expected interactions with that student. Adding two Black students to the class would generate another 1.4 emails with Black students for every White student. Thus, the negative effects of subtracting two Black students might have a greater effect on total Black-White interactions than the negative effect of complete housing segregation discussed above.

III.B. Does A "Group Means" Model Approximate Social Networks?

As discussed above, most peer effects studies assume that an agent is affected by the mean characteristic or outcome of the other individuals in the group. Studies of peer effects in primary school often take the classroom as the relevant group and university level studies may take the cohort-major cell or the freshman dormitory as the relevant group. Here we have the opportunity to ask how much students actually interact with the other students assigned to their group. We then calculate mean SAT scores and Dartmouth GPAs for each student's dorm group and for their *actual* peer group campus wide as demonstrated by whom they email. We show the correlation between mean test scores (and GPA) for the "actual" group and their freshman dorm group.

For the freshmen (class of 2006) about 19 percent of total emails are exchanged with members of their freshman dorm. The sophomores exchange 16 percent of total emails with classmates from their freshman dorm, and this percentage falls to 8 percent for the seniors. If instead of weighting by email volume (as above) we ask what fraction of correspondents come from one's freshman dorm, we get roughly the same answer.

Perhaps the more relevant question for peer effects research is how correlated "actual" peer ability is with constructed group ability when groups are formed by the econometrician around freshman hallway assignments. To address this, we calculate for each student average peer SAT scores (and GPA) in four ways. First, we weight peer SATs (GPA) by the volume of email exchanged with that peer. If zero emails are exchanged, the peer observation is given zero weight. Second, we calculate mean peer SATs using the simple average over all other students on one's freshman hallway, excluding own observation. Third, we calculate the mean within one's *hallway and race* cell. Finally, we calculate mean peer SAT weighting by the predicted email volume with that peer. We predict volume using the set of regressions in Table VI which include all student characteristics for both students in a pair. This last measure is intended to be an extreme (though unrealistic) upper bound on what a researcher could predict about peer ability given all of the right-hand-side variables.

The hallway mean SAT score is relatively uncorrelated with the actual (emails) weighted mean. The correlation is .03. However, the dorm-race mean has a correlation of .29 with the actual weighted mean. The prediction weighted mean of peer SATs has a correlation of .42 with the actual weighted mean.

The message is that race matters in the formation of peer groups. Even among a relatively homogeneous group of college students (relative to the set of all U. S. college students), race plays a large role in determining social interactions. If researchers are seeking to form the most realistic peer groups, one specification to include is one in which peer groups condition on race. Conditioning on all the other right-hand-side variables (gender, fraternity membership, financial aid, status, etc.) AND weighting these variables by the regression coefficients raises the correlation from .29 to .42.

We find a similar message when we examine peer Dartmouth GPAs rather than peer SATs. Mean GPA of freshman hallway group has a .08 correlation with actual weighted GPA. Mean GPA for the dorm-race cell has a correlation of .23 with actual weighted GPA.

In Figure II we examine what fraction of Black students' emails are exchanged with peers in various deciles of the SAT distribution. The horizontal axis shows the decile of the Dartmouth SAT distribution. The "Actual" line shows that roughly 26 percent of emails were exchanged with students in the first decile. The relatively flat line ("Geography") shows that an equal fraction of emails would fall in each decile if the students' interactions were based purely on geography (freshman dorm and room location). We simulated this distribution by predicting emails for every pair of students using a Poisson regression which included only the geographic variables on the right-hand side. We then calculated what fraction of predicted emails fell in each decile.

The "Race and Geography" line shows what happens when Black students interact based on geography and race. These are calculated from predicted number of emails where we predict using a Poisson regression of emails on the race dummies for person 2, geographic variables, and the interactions of race and same dorm. Here we see that the addition of the race information yields a predicted SAT distribution that is close to the actual distribution. When we predict using all the right-hand-side vari-

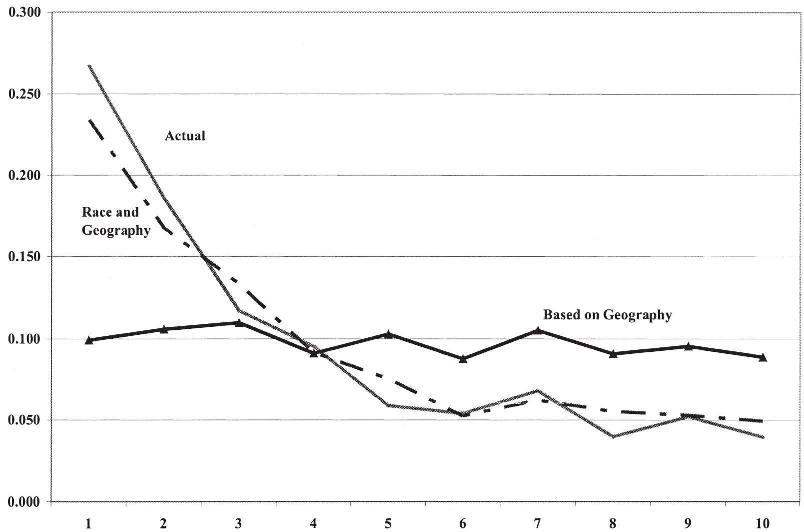

FIGURE II

Actual and Hypothetical SAT Distribution of Black Students' Peers Weighting by the Level of Interaction between Each Student and Her Peers

This shows the fraction of Black students' emails that are exchanged with students in each decile of the Dartmouth SAT distribution. The line labeled actual shows the true distribution. The flat line labeled "geography" shows what the distribution would look like if students used only freshman year location (geography) to determine their peer group. This uses predicted email volumes from a Poisson regression of email volumes on all of the geographic variables. The "race and geography" line shows the distribution that results if students use both the race and geography to determine their peer group. If we predict email volumes using all Xs for each student in the pair (not shown on the graph), we get a peer SAT distribution that matches the Actual line almost exactly.

ables, the predicted line lies almost directly on top of the actual line (not shown).

III.C. How Much Does Using a Better Peer Group Proxy Affect Estimated Peer Effects?

Finally, we ask how much measured peer effects change if we switch from a dorm mean to a mean that weights based by the level of interaction (emails). In Table VIII we run linear probability models of own decision to join a fraternity or a sorority on the peer average decision. We focus on this outcome rather than on an academic outcome like GPA because previous work (e.g., Sacerdote [2001], Zimmerman [2003], Duncan et al. [2003b], and Foster [2004]) indicates that peer effects in social outcomes are larger and more robust than peer effects in academic outcomes.

TABLE VIII

PEER EFFECTS IN FRATERNITY MEMBERSHIP

We estimate peer effects in the decision to join a Greek organization using four different specifications. In column (1) we run own decision on the average of the dormmates' decision as in Sacerdote [2001]. In column (2) we use the average decision of all peers where we weight the peer outcome by the amount of email volume exchanged with that peer. In column (3) we follow Foster [2003] by instrumenting for the dorm mean of fraternity membership using average characteristics of randomly assigned dormmates. Instruments are dormmate means of the following variables: combined SAT score, Academic index, NY exam school status, Prep school status, and Financial aid status. First-stage regressions are shown in Appendix 2. In column (4) we use the same set of instruments to instrument for the email weighted peer average of fraternity membership.

	(1) Greek member (OLS)	(2) Greek member (OLS)	(3) Greek member (IV)	(4) Greek member (IV)
Freshman dorm mean of "Greek"	0.711		0.505	
	(0.069)**		(0.196)*	
Email weighted peer mean of "Greek"		0.990		0.771
		(0.037)**		(0.294)**
Student is Black	−0.059	0.036	−0.064	0.012
	(0.071)	(0.058)	(0.071)	(0.068)
Student is Asian	−0.005	0.047	−0.011	0.031
	(0.040)	(0.033)	(0.041)	(0.040)
Student is Hispanic	0.052	0.068	0.051	0.064
	(0.059)	(0.049)	(0.060)	(0.050)
Student is other non-White	−0.038	−0.032	−0.037	−0.032
	(0.096)	(0.079)	(0.096)	(0.080)
Varsity athlete (0–1)	0.092	0.024	0.097	0.044
	(0.033)**	(0.027)	(0.033)**	(0.037)
Sat Score (math + verbal)	−0.000	−0.000	−0.000	−0.000
	(0.000)*	(0.000)	(0.000)*	(0.000)
Male	0.023	0.038	0.020	0.032
	(0.027)	(0.023)	(0.028)	(0.024)
Went to NYC exam school	0.043	0.009	0.042	0.015
	(0.081)	(0.067)	(0.081)	(0.068)
Went to fancy prep school	0.101	0.041	0.111	0.062
	(0.061)	(0.051)	(0.062)	(0.059)
Receives financial aid (0–1)	−0.136	−0.063	−0.140	−0.083
	(0.028)**	(0.023)**	(0.028)**	(0.035)*
Constant	0.704	0.213	0.803	0.393
	(0.232)**	(0.192)	(0.249)**	(0.315)
Observations	1233	1233	1233	1233
R^2	0.122	0.399	0.115	0.382

Standard errors are in parentheses.
* Significant at 5 percent; ** significant at 1 percent.

In column (1) we regress a student's decision to join a fraternity on the dorm mean for the same outcome and find a coefficient of .71, with a t-statistic of 10.3. If we instead weight the peer mean outcome by emails exchanged between the student and the other students on campus, the peer effect coefficient increases to 1.0 (column (2)).

This latter coefficient is of course biased by the fact that students sort into their peer groups, and we are not taking advantage of the randomized assignment of freshman housing. As a partial solution, we follow Foster [2003] and instrument for the peer average of fraternity status using the average background characteristics for the randomly assigned peers in a student's freshman hallway group. Our instruments for peer group mean outcomes are the dorm averages of SAT scores, academic index, New York exam high school attendance, prep school attendance, and financial aid status. The first-stage regressions are shown in Appendix 2. In column (1) of Appendix 2 the dependent variable is the dorm average fraternity status, and the instruments predict this outcome with an F statistic of 13.0. In column (2) the dependent variable is the email weighted average of fraternity status, and the F statistic is 7.4.

In Table VIII column (3) we regress own fraternity status on the hallway mean of fraternity status where the instruments are the hallway mean background characteristics described above. The peer effect coefficient is .51 and is statistically significant. In column (4) we regress own outcome of fraternity status on peer average outcome, where the peer outcome is weighted by email volumes. We again instrument for peer outcomes using hallway average characteristics. Using our emails to define peer groups, we find a peer effect that is 50 percent larger than that found using hallways to define peer groups. However, given the lack of precision, we cannot reject that the peer effects coefficients in columns (3) and (4) are equal.

The message from this exercise is a fairly intuitive one: conventional peer effects estimates may understate the total peer influences experienced by an individual because researchers rarely know the true peer group and must therefore form peer groups based on observables like classroom, or dorm assignment or neighborhood. In our case, the estimated effect increases by about 50 percent when we use email volumes to first determine who is interacting with whom before we estimate the peer effect.

IV. Conclusion

We find that geographic closeness, racial similarity, family background, and common interests like academic majors, Greek organizations, and varsity athletics all have positive effects on the likelihood that two students interact. Cross-race interactions are much less likely than within-race interactions. Two White students in the same class have about a 2 percent chance of interacting whereas a White and a Black student in the same class have a .8 percent chance of interacting. Placing either of these pairs in the same dorm multiplies the likelihood of interaction by a factor of 3. Expressed in terms of email volume, randomly chosen White-White pairs send each other about .7 emails versus .2 emails for White-Black pairs. Putting either pair in the same dorm raises the volume of email between the two by a factor of 3 or 4.

There does not appear to be any extended network effect in which having a Black roommate increases a White student's interaction with other Black students outside of her room, hallway, or dorm. When we simulate different housing policies, it becomes clear that it is difficult to create meaningful amounts of additional interracial interaction simply by moving students around. This is in part because the proximity effect is relevant for very small distances, so it is impossible to make a given student closer to large numbers of other students. The vast majority of a student's interracial interactions take place outside her freshman dorm group simply because only 1 percent of other students on campus share her same freshman dorm.

In contrast, the effects of being in the same entering class on the amount of interaction are also large in magnitude and operate on a much larger group (roughly 1050 students). Thus, changes in the racial or SAT makeup of my class could have a large effect on the characteristics of my peer group. Within the context of this study, differences in ability across students do not create a sizable barrier to interaction. A low SAT scoring student is almost as likely to interact with his classmates and dormmates as any other student.

These facts help explain why colleges and universities go to great lengths to manage the composition (racial, geographic, athletic) of each incoming freshman class. Adding a few more international students to a class can have a large effect on the likelihood that the modal student in that class interacts with one of these international students.

We asked whether a linear in means approach is a good approximation to constructing a student's true peer group. We find that the mean ability (SAT or GPA) of one's hallmates is only modestly correlated with the mean ability of one's correspondents as weighted by email volume. However, if we use the hallway-race cell, we get a much stronger correlation between the group mean and the mean ability of actual peers as identified through email exchanges.

Overall, the study indicates that small differences in location can have large impacts in the amount that the two people interact. We posit that this is because small distance costs matter when the other person and hence the benefit of the interaction is unknown. Race appears to be a real barrier to interaction, but the positive effects of proximity on interaction can offset the negative effects of two people being from a different race.

Proximity, race, family background, and interests all determine who interacts with whom. And these interactions blossom into friendships which can have a profound influence on our lives, our career choices, and perhaps our preferences.

APPENDIX 1: COMPARISON OF PRIMARY AND SECONDARY USERS

The Primary users are those who participated in the study and for whom we have a complete record of their outgoing and incoming email volumes. The Secondary users are people who appear in the data set by virtue of having been in contact with one of the Primary users. The set of Primary users is large enough that virtually every student (and most faculty members) appear in the data set at least once. The comparison in the table is a simple t-test of the difference in means for the difference between the two groups.

Variable	Obs Primary = 0	Obs Primary = 1	Mean Primary = 0	Mean Primary = 1	T-test for difference in means	T-test difference for juniors & seniors
Male	2,684	1,250	0.51	0.48	1.423	1.207
Member fraternity/ sorority	2,684	1,250	0.33	0.49	−9.749	−1.422
White	2,463	1,250	0.60	0.72	−2.561	−0.157
Black	2,463	1,250	0.07	0.05	2.957	2.350
Asian	2,463	1,250	0.13	0.14	−0.881	−2.821
Hispanic	2,463	1,250	0.07	0.06	0.885	1.196
Academic index (from admissions)	2,624	1,250	212.12	215.30	−6.412	−4.915
Cumulative GPA	2,684	1,250	3.28	3.38	−8.044	−6.479
Math SAT score	2,609	1,250	706.30	714.738	−4.239	−3.628

Appendix 2: First-Stage Regressions for Table VIII

Table VIII columns (3) and (4) use average background characteristics of randomly assigned dormmates to instrument for the freshman dorm mean of fraternity status ("Greek") and the peer mean of fraternity status where the peer average is weighted by email volumes.

	(1) Freshman dorm mean of "Greek"	(2) Peer mean of "Greek" using email volumes as weights
Dorm average SATs	0.002	0.001
	(0.000)**	(0.001)
Dorm average academic index	−0.014	−0.006
	(0.003)**	(0.005)
Dorm average of NYC exam school status	−0.617	−0.640
	(0.194)**	(0.319)*
Dorm average for prep school status	−0.810	−0.005
	(0.129)**	(0.212)
Dorm average for financial aid status	−0.601	−0.302
	(0.057)**	(0.094)**
Black (0–1)	−0.009	−0.111
	(0.028)	(0.045)*
Asian (0–1)	−0.026	−0.071
	(0.016)	(0.026)**
Hispanic (0–1)	−0.003	−0.019
	(0.023)	(0.038)
Other non-White (0–1)	−0.011	−0.013
	(0.037)	(0.061)
Varsity athlete (0–1)	0.016	0.079
	(0.013)	(0.021)**
Combined SAT score	0.000	−0.000
	(0.000)	(0.000)*
Male (0–1)	−0.012	−0.022
	(0.011)	(0.017)
Went to NYC exam school (0–1)	−0.005	0.033
	(0.031)	(0.052)
Went to fancy prep school (0–1)	0.035	0.093
	(0.024)	(0.039)*
Financial aid status (0–1)	−0.011	−0.081
	(0.011)	(0.018)**
Constant	1.780	1.131
	(0.371)**	(0.610)
Observations	1233	1233
R^2	0.138	0.083
F(15, 1217) =	13.02	7.37

Standard errors are in parentheses.
* Significant at 5 percent; ** significant at 1 percent.

Dartmouth College

References

Abu-Ghazzeh, Tawfiq M., "Housing Layout, Social Interaction, and the Place of Contact in Abu-Nuseir, Jordan," *Journal of Environmental Psychology,* XIX (1999), 41–73.

Alesina, Alberto, Reza Baqir, and Caroline M. Hoxby, "Political Jurisdictions in Heterogeneous Communities," *Journal of Political Economy,* CXII (2004), 348–396.

Amicus Brief of Harvard University, Brown University, University of Chicago, Dartmouth College, Duke University, the University of Pennsylvania, Princeton University, and Yale University, Case Numbers 02-241 and 02-516, Barbara Grutter v. Lee Bollinger, et al., Supreme Court of the United States, 2003.

Arcidiacono, Peter, Gigi Foster, and Joshua Kinsler, "Estimating Spillovers in the Classroom with Panel Data," mimeo, Duke University, 2004.

Betts, Julian R., and Andrew Zau, "Peer Groups and Academic Achievement: Panel Evidence from Administrative Data," Working Paper from University of California, San Diego and Public Policy Institute of California, 2004.

Blau, Peter Michael, *Exchange and Power in Social Life* (New York: Wiley, 1964).

Bowen, William G., and Derek Bok, *The Shape Of The River: Long-Term Consequences Of Considering Race In College And University Admissions* (Princeton, NJ: Princeton University Press, 1998).

Bowen, William G., and Sarah A. Levin, *Reclaiming The Game: College Sports And Educational Values* (Princeton, NJ: Princeton University Press, 2003).

Case, Anne C., and Lawrence F. Katz, "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youths," National Bureau of Economic Research Working Paper No. 3705, 1991.

Duncan, Greg, Johanne Boisjoly, Michael Kremer, Daniel Levy, and Jacquelynne S. Eccles, "Empathy or Antipathy? The Consequences of Racially and Socially Diverse Peers on Attitudes and Behaviors," Joint Center for Poverty Research Working Paper No. 326, February 2003a.

Duncan, Greg, Johanne Boisjoly, Michael Kremer, Daniel Levy, and Jacquelynne S. Eccles, "Peer Effects in Binge Drinking among College Students," mimeo, Northwestern University, December 2003b.

Epstein, Richard, "A Rational Basis for Affirmative Action: A Shaky But Classical Liberal Defense," *Michigan Law Review,* C (2002), 2036–2053.

Evans, William N., Wallace E. Oates, and Robert M. Schwab, "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy,* C (1992), 966–991.

Falk, Armin, and Michael Kosfield, "It's All about Connections: Evidence on Network Formation," mimeo, University of Zurich, 2003.

Festinger, Leon, Stanley Schachter, and Kurt Back, *Social Pressures In Informal Groups: A Study Of Human Factors In Housing* (Stanford, CA: Stanford University Press, 1963).

Foster, Jennifer, "Peerless Performers," mimeo, University of South Australia, 2003.

——, "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism," mimeo, University of South Australia, 2004.

Gaspar, Jess, and Edward Glaeser, "Information Technology and the Future of Cities," *Journal of Urban Economics,* XLIII (1998), 136–156.

Glaeser, Edward, and Bruce Sacerdote, "The Social Consequences of Housing," *Journal of Housing Economics,* IX (2000), 1–23.

Graham, Bryan S., "Essays on the Econometrics of Social Interactions through Excess Variance Contrasts," Ph.D. thesis, Harvard University, 2005.

Graham, Bryan S., and Jinyong Hahn, "Identification and Estimation of the Linear-in-Means Model of Social Interactions," *Economics Letters,* LXXXVIII (2005), 1–6.

Harris, Judith Rich, *The Nurture Assumption: Why Children Turn out the Way They Do* (New York, NY: The Free Press, 1998).

Hills, Lindsay Marie, "A Thin Line between Red and Green: Emerging Patterns in

Group Preference an Exploration at Bryn Mawr College," Bryn Mawr Senior Thesis, 2003.

Holahan, C. J., B. L. Wilcox, and M. A. Burnam, "Social Satisfaction and Friendship Formation as a Function of Floor Level in High-Rise Student Housing," *Journal of Applied Psychology,* LXIII (1978), 527–529.

Hoxby, Caroline M., "Peer Effects in the Classroom: Learning from Gender and Race Variation," National Bureau of Economic Research Working Paper No. 7866, 2000a.

——, "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review,* XL (2000b), 1209–1238.

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

Marmaros, David, and Bruce Sacerdote, "Peer Effects in Occupational Choice," *European Economic Review,* XLVI (2002), 870–879.

Rainio, K., "A Study of Sociometric Group Structure: An Application of a Stochastic Theory of Social Interaction," in J. Berger, M. Zelditch, and B. Anderson, *Sociological Theories in Progress,* Vol. 1 (Boston, MA: Houghton Mifflin, 1966).

Reid, Stuart A., "Students Question True Diversity of Campus," *The Dartmouth,* CLXII (February 4, 2005), 1.

Richards, Phillip, "Prestigious Colleges Ignore the Inadequate Intellectual Achievement of Black Students," *The Chronicle of Higher Education,* XLIX (September 13, 2002), B11.

Richeson, Jennifer A., Abigail A. Baird, Heather L. Gordon, Todd F. Heatherton, Carrie L. Wyland, Sophie Trawalter, and J. Nicole Shelton, "An Fmri Investigation of the Impact of Interracial Contact on Executive Function," *Nature Neuroscience,* VI (2003), 1323–1328.

Sacerdote, Bruce I., "Peer Effects with Random Assignment," *Quarterly Journal of Economics,* CXVI (2001), 681–704.

Schapiro, Amy, "Unlocking Emory's Racial Segregation," *The Emory Wheel,* LXXXV (April 4, 2003).

Stinebrickner, Todd, and Ralph Stinebrickner, "What Can Be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students From Disadvantaged Backgrounds," mimeo, 2004.

Tuma, Nancy B., and Maureen T. Hallinan, "The Effects of Sex, Race and Achievement on Schoolchildren's Friendships," *Social Forces,* LVII (1979), 1265–1285.

Waller, Willard, *The Family: A Dynamic Interpretation* (New York, NY: Dryden, 1938).

Winston, Gordon C., "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education," *Journal of Economic Perspectives,* XIII (1999), 13–36.

Zimmerman, David J., "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics,* LXXXV (2003), 9–23.