

# HOW DO FRIENDSHIPS FORM?

David Marmaros

*Google.com*

Bruce Sacerdote\*

*Dartmouth College and NBER*

bruce.sacerdote@dartmouth.edu

December 23, 2004

## Abstract

We examine how people form social networks among their peers. We do this by using a unique dataset 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. The data are consistent with a model in which the expected value of a social interaction with an unknown person is low relative to travel costs and the benefits from interacting with the same person repeatedly are high. First year students interact with whomever is in their immediate proximity and form long term friendships with a subset of those people. Geographic proximity and race are even 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. We show that a traditional "linear in means" model of social interaction can be a good approximation to the actual set of social interactions if researchers condition on key factors like distance, race and age.

---

\* 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 Wharton, DELTA, and the NBER for helpful suggestions. JEL Classification Codes: I2, J0. Keywords: Social Networks, Peer Effects.

## 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. (See for example Harris [1998], Case and Katz [1991], Evans Oates and Schwab [1992], Zimmerman [2003], Hoxby [2000a], and Marmaros and Sacerdote [2002]). 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 by on how we actually choose and are chosen by a specific group of friends within a neighborhood, school or workplace.<sup>1</sup> 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 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].<sup>2</sup> 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 provides a very direct measure of the amount of racial segregation on a campus. Bowen and Bok [2001] 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

---

<sup>1</sup> One exception is Falk and Kosfeld [2003] who study the shape of networks in an experimental setting.

<sup>2</sup> 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.

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, Harvard, Yale, Brown, University of Chicago, Duke, University of Pennsylvania and Princeton) 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 [2001], 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 [2002] 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 Shapiro [2003] describing Emory University or Hills [2003] describing Bryn Mawr.

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

Proximity does have a large effect (3x) on the likelihood of social interaction among individuals of different races. 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 (ie within building.) Given the physical reality that a student can be housed in truly close proximity to 50 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.<sup>3</sup>

---

<sup>3</sup> For an average white student only about 9 percent of her interactions with black students from her freshman dorm. This means that these same dorm interactions are much more likely than out of dorm interactions but still a small fraction of overall cross race interactions.

In contrast, placing two students in the same entering class (cohort) has a 6x 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 incredibly important in determining my 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.

The majority of existing peer effects studies use a linear in means approach to creating measures of peer background ability or peer outcomes. Researchers typically use the mean outcome or mean pre-treatment characteristic for a group which is assumed to represent the individual's peers or friends. For some examples, see Graham [2004], Betts [2004], Hoxby [2002] and the studies referenced above. The econometrics of social interactions literature including Manski [1993], Graham and Hahn [2004] 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 [2001], Duncan et al [2003], Foster [2002, 2003], Arcidiacono et al [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. Group means are highly correlated with means for actual peers' outcomes if we construct peer groups along lines that truly matter, like race, entering class, and geographic distance.

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.

We ask whether having a minority roommate or dormmate introduces white students into a broader network of minority students. We take various subgroups of students (e.g. the white students, the black students, the athletes) and examine the distribution of academic ability for their peers. We ask whether Dartmouth could further increase interracial interaction simply by rearranging freshman housing assignments.

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 also allows one test of the Gaspar Glaeser thesis [1998] that email communication is a complement to face to face communication, rather than a substitute.

#### *On Peers, Race and Location*

There is a burgeoning literature on peer effects at the elementary, secondary, and post-secondary 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 [2002], Kremer and Levy [2000], Foster [2002], 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 et al [1963], Abu-Ghazzeh [1999], and Holahan, Wilcox and Burnam [1978] have emphasized the importance of geographic proximity in determining who interacts with whom. Festinger gathered data on social interactions among new

MIT 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 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 [1978] 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.

#### *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.<sup>4</sup>

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

---

<sup>4</sup> 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.

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 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:

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

where  $g(.)$  is a function that increases with the number of previous social interactions and  $c(.)$  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 [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.

---

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.

Suppose that interacting within a student of a different race is more costly than interacting within race.<sup>5</sup> 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 new 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.

### *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.<sup>6</sup> 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  $\mathbf{X}\beta = \alpha + \mathbf{B1}^*(\text{dummies for person 1's race, varsity athlete status, gender, Greek status, type of high school, financial aid status}) + \mathbf{B2}^*(\text{dummies for person 2's race, varsity athlete status, gender, Greek status, type of high school, financial aid status}) + \beta_3^*(\text{dummy for same graduating class}) +$

---

<sup>5</sup> Indeed the neuroscience literature suggests that white-black interaction is more stressful than within race interaction and that this effect can be measured in functional magnetic resonance imaging (fMRI) scans (Richeson et. al. [2003]).

<sup>6</sup> 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.

$\beta_4$ \*(dummy for same freshman dorm) +  $\beta_5$ \*(interactions of race and same freshman dorm) +  $\beta_6$ \*(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 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 13 months long). Thus the number of emails should have a Poisson distribution. Second, in the data, the number of emails appears to have a Poisson distribution. 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.<sup>7</sup>

In our regression tables we report the regression coefficients, standard errors and  $e^{\text{coefficient}}$ . 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  $e^{x\beta}$ , increasing  $x$  by one multiplies the predicted value by  $e^\beta$ ). For example, coefficient on "same freshman dorm" is estimated to be in the range of 1.3 which means that being in the same freshman dorm raises the number of emails by a factor of 3.7.<sup>8</sup>

### *Do Emails Equal Friendship?*

---

<sup>7</sup> We are greatly indebted to both referees for suggesting that we switch from OLS to Poisson.

<sup>8</sup> 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.

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 intra-campus email messages instantaneously, 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 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 dataset. 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 10 or more emails are sent between A and B during the 13 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 10, and the average volume is .21 messages. In short, we observe a huge 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 towards 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.<sup>9</sup>

## **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 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 message 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 in to 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.<sup>10</sup> 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

---

<sup>9</sup> 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.

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.<sup>11</sup>

Table 1 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 13 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 2 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 2 also shows averages for several measures of academic ability including incoming math SAT score and incoming Academic Index. The Academic Index can range from 60 to 240 and is a weighted average of SAT I scores (weight=1/3), SAT II scores (weight=1/3), and re-scaled

---

<sup>10</sup> In other words, we capture not just volumes sent during that particular session, but any information in the student's folders.

high school class rank (weight=1/3). 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.<sup>12</sup> Five percent of primary users attended one of the well known, selective private schools that serves as a major feeder high school for Dartmouth.<sup>13</sup> 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.

A natural question to ask is whether the primary users are representative of Dartmouth students as a whole. We address this question in Appendix Table 1 where we compare the primary users to 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.

---

<sup>11</sup> Eight percent of NetBlitz users opted out of the study. NOTE TO REFEREES: We are working on an analysis of who opted out and will have this for the final version.

<sup>12</sup> 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.

<sup>13</sup> We created a dummy equal to 1 for students from private prep schools that fall within the top 20 feeder 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.

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 may still be well positioned to study cross race interaction because the only emails that do not enter our dataset 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 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 have emails going in both directions during the sample period.

Table 3 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.<sup>14</sup> In .2 percent of pairs 5 or more emails have traveled in each direction (for a total volume of 10 or more).<sup>15</sup> 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 range is enormous. Roughly 24 percent of the pairs consist of two members from the same

---

<sup>14</sup> We define an indicator variable called "talks" which equals 1 if persons 1 and 2 have sent and received emails from each other.

<sup>15</sup> Our empirical results are similar whether our dependent variable is the actual volume, or a dummy for 2 or more emails sent (total) or a dummy for 10 emails sent (total).

graduating class. .3 percent of pairs 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.<sup>16</sup>

### III. Results

In Table 4 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 in which person 1 (the primary user) is white. We look at the amount of interaction based on whether person 2 is black (yes or no) and in the same freshman dorm as person 1 (yes or no).<sup>17</sup> The top row shows that white-non-black pairs that are not in the same freshman dorm, exchange .71 emails on average. If both people are in the same dorm and class, the mean jumps to 2.95 indicating a multiplicative effect from same dorm of 4.2x.

When person 1 is white and person 2 is black, the multiplicative effect of being in the same dorm is 2.93, 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 during the 13 month sample period. If these two students were in the same dorm and class, this mean rises to .62.

Now we consider the race effect. Holding constant that both students are in the same dorm and class, the second student being black reduces mean email volume by a factor of 4.76 (i.e.

---

<sup>16</sup> 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.

2.95/.62). For students not in the same class and dorm, the second student being black reduces email volume by a factor of 3.34.

We also show the mean probability that the two people talk at all, ie exchange at least one email in each direction. 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 3-4x 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 3x effect whether we consider interaction within versus across freshman dorms.

In Tables 5 and 6 we analyze the data at the student level rather than the pair level. We ask how much having a black freshman roommate or above median percentage of black students on one's freshman hallway increases a student's interaction with black students. The first panel of Table 5 cuts the data by own race (black versus not) and a dummy for having any black roommate<sup>18</sup>. We report students' percent of email volume that is exchanged with black students and the percent of their correspondents who are black. For all the non-black students, having a black

---

<sup>17</sup> In results not reported, we limited the sample to white-black plus white-white pairs and found similar means.

<sup>18</sup> If a white student has two roommates, the dummy is set to 1 if either of her roommates is black.

roommate raises the fraction of emails volume exchanged with black students by .3 percent. This only amounts to an extra 3 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.

In Table 6 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 one's dorm, one's hallway, and a dummy for having black roommate.<sup>19</sup> We control for all observed student characteristics including gender, SAT scores, financial aid status, type of high school attended, Greek and athletic status. 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. 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 dorm.<sup>20</sup>

In Table 7 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  $e^{\text{coefficient}}$  in square brackets []. The purpose of the latter is to show the multiplicative effect of a one unit change in the right hand side variable. In this table the omitted categories for person 1 and 2's race is always white. The interaction dummies are constructed to be multiplied together to calculate the

---

<sup>19</sup> The dorm and hallway percent black calculation excludes the student herself.

total effect. In other words, the interaction effect for person 2 being "black and in the same class and dorm" is in addition to 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 dummies 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 (for same dorm and person 2 black) that we observed in the means of the raw data in Table 4. The regression shows that being in the same freshman dorm (holding same class constant) multiplies the number of emails by 4.13. Table 4 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). Table 4 indicates the same percent reduction in email volume.

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. First, 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<sup>21</sup>. 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

---

<sup>20</sup> A 10 percent increase in the fraction black would correspond to adding about 3 more black students to the white student's dorm. Using the means for same dorm versus not in Table 4, we would expect an additional  $.4*3=1.2$  messages.

<sup>21</sup> "Same freshman dorm" always implies same class.

.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. 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.<sup>22</sup> 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 status, Greek status and majors are all determined endogenously, we offer these as descriptive statistics rather than causal effects. Furthermore, we don't 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 pre-treatment 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. We ran the fully interacted specification and 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 9x 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 non-aid – non-aid. 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 7 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,

---

<sup>22</sup> Here we multiplied the three relevant effects from person 1 is an athlete, person 2 is an athlete and the interaction

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 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.<sup>23</sup> The coefficients on the terms interacting person 2's race with the race of 1's roommates and the percent black in 1's dorm are not distinguishable from zero. This is consistent with the student level analysis we showed in Table 6.

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.

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 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 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 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

---

term to get the total effect.

<sup>23</sup> There is still of course the direct effect that being close to any student of any race makes 1 more likely to interact with that particular student.

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 and white-black pairs have a slightly smaller effect. There are still 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 8 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 7 (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 2.3x effect over and above the 2.0x effect from being in the same dorm, for a total effect of 4.6x. Being in the same room adds an additional effect of 3.1x. 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.<sup>24</sup>

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)

---

<sup>24</sup> The regression controls but does not report the effect from the two students being in the same class.

that this fact limits the degree to which proximity can be used by policy makers to create shifts in the overall amount inter-group interaction that occurs.

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

#### *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. In Table 9 we report 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 have 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 7 column (2). We reassign black freshmen across dorms to deliver a uniform percent black in each dorm, for a given class year. This is possible because the actual random assignment to groups of 30 to 50 yielded some dorms with 0 or even 6 black students rather than the roughly 3 of a perfectly even distribution. Across the entire campus, we predict an additional 146 emails between white - black pairs. This indicates that each white student would exchange an additional .16 emails with black students, or a less than one percent increase on the mean of 18-19 emails currently exchanged with black students during the sample period.

The intuition for this finding is as follows: We found in Tables 4 and 7 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 4. 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  $3 \times .4 = 1.2$  emails exchanged with black students. This is again modest relative to the current average of 19 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 large effect on my social interactions. A student interacts 5-7x 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 1,050 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.

#### *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.

In Table 10 we show the fraction of total emails that are exchanged between a student and her 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 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 group ability when groups are formed around freshman dorm assignments. In Table 11, 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. (For example, 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 in one's freshman dorm group, excluding own observation. Third we calculate the mean within one's *dorm 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 7 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 or outcomes given all the right hand side variables.

The upper panel of Table 11 shows the correlation between actual peers mean SAT and the three possible peer group means: dorm mean, dorm-race mean, and mean weighting by predicted interactions. The dorm mean is almost uncorrelated with the actual weighted mean. The correlation is .006. However the dorm-race mean has a correlation of .35 with the actual weighted mean. The prediction weighted mean has a correlation of .51 with the actual weighted mean.

The message is that race matters in the formation of peer groups. Even among a relatively homogenous group of college students (relative to the set of all US college students), race places 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 .35 to .51.

We find a similar message when we examine peers Dartmouth GPAs rather than peer SATs. Mean GPA of freshman dorm group has a .10 correlation with actual weighted GPA. Mean GPA for the dorm-race cell has a correlation of .23 with actual weighted GPA.

In Figure 2 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 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 an predicted SAT distribution that is pretty close to the actual distribution. When we predict using all the right hand side variables, the predicted line lies almost directly on top of the actual line.

### *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 12 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 [2003], Foster [2003]) indicates that peer effects in social outcomes are larger and more robust than peer effects in academic outcomes.

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 .49, with a t-statistic of 8.8. If we instead weight the peer mean outcome by emails exchanged between the student and the other students on campus, the peer effect coefficient doubles 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 email weighted peer average using the average of the randomly assigned freshman dormmates. This instrumental variables regression yields a peer effect of 1.4.

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

## **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. 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 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 sizeable 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 (for example) 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 dormmates is only modestly correlated with the mean ability of one's correspondents (weighted by email volume). However, if we use the dorm-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 our preferences.

## *References*

- Abu-Ghazze, Tawfiq M., "Housing Layout, Social Interaction, and the Place of Contact in Abu-Nuseir, Jordan," *Journal of Environmental Psychology*, Vol. 19, No.1, March 1999, pp. 41-73.
- Alesina, Alberto, Reza Baqir and Caroline Hoxby, "Political Jurisdictions in Heterogeneous Communities" National Bureau of Economic Research Working Paper # 7859, 2000 .
- 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 Josh 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, P.M., *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, N.J. : Princeton University Press, 1998.
- Bowen, William G. and Sarah A. Levin, *Reclaiming The Game : College Sports And Educational Values*, Princeton, N.J. : 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 # 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. #326 February 2003.
- Duncan, Greg, Johanne Boisjoly, Michael Kremer, Daniel Levy and Jacquelynne S. Eccles "Peer Effects in Binge Drinking Among College Students," Mimeo, Northwestern University, December 2003.
- Epstein, Richard, "A Rational Basis for Affirmative Action: A Shaky But Classical Liberal Defense," *Michigan Law Review*, 100 (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 Bonn and University of Zurich, 2003.

- Festinger, Leon and Stanley Schachter and Kurt Back, *Social pressures in informal groups : a study of human factors in housing*, Stanford: Stanford University Press, 1963.
- Foster, Jennifer, "Peerless Performers," Mimeo 2003, University of South Australia.
- Foster, Jennifer, "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism," Mimeo 2004, University of South Australia.
- Gaspar Jess and Edward Glaeser, "Information Technology and the Future of Cities," *Journal of Urban Economics*, Vol. 43, no. 1 (January 1998): 136-156.
- Gaviria, Alejandro and Stephen Raphael, "School Based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, v.83 (2001), 257-68.
- Glaeser, Edward and Bruce Sacerdote, "The Social Consequences of Housing," *Journal of Housing Economics*, vol. 9, 2000.
- Graham, Bryan S. and Jinyong Hahn "Identification And Estimation Of The Linear-In-Means Model Of Social Interactions," mimeo Harvard University and University of Southern California, 2003.
- Graham, Bryan S. "Identifying Social Interactions Through Excess Variance Contrasts," Ph.D. Dissertation, Harvard University, 2004.
- Grinblatt, Mark, Matti Keloharju and Seppo Ikaheimo, "Interpersonal Effects in Consumption: Evidence from the Automobile Purchases of Neighbors," National Bureau of Economic Research Working Paper #10226, 2004.
- Hallinan, Maureen T., "The Process of Friendship Formation," *Social Networks*, vol 1., 1978/9, pp. 193-210.
- Harris, Judith Rich, *The Nurture Assumption: Why Children Turn out the Way They Do*, The Free Press, NY, 1998.
- 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*, v. 63, no. 4, 1978, pp. 527-529.
- Hoxby, Caroline M., "Peer Effects in the Classroom: Learning from Gender and Race Variation," National Bureau of Economics Working Paper Number 7866, 2000.
- Hoxby, Caroline M, "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, vol. 90, no. 5, December 2000, pp. 1209-38
- Hanushek, Eric A., John F. Kain, Jacob M. Markman, Steven G. Rivkin, "Does Peer Ability Affect Student Achievement?" National Bureau of Economics Working Paper Number 8502, 2001.
- Hills, Lindsay Marie, *A Thin Line Between Red and Green: Emerging Patterns in Group Preference an Exploration at Bryn Mawr College*, Senior Thesis, 2003.

- Kremer, Michael and Daniel Levy "Peer Effects from Alcohol Use Among College Students," mimeo, 2000.
- Kremer, Michael, "How Much Does Sorting Increase Inequality?," *Quarterly Journal of Economics*, CXII (1997), 115-139.
- 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*, 46 (May 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: Houghton Mifflin, 1966.
- Richards, Phillip, "Prestigious Colleges Ignore the Inadequate Intellectual Achievement of Black Students," *The Chronicle of Higher Education*, September 13, 2002.
- Richeson, Jennifer A, Abigail A Baird, Heather L Gordon, Todd F Heatherton, Carrie L Wyland, Sophie Trawalter & J Nicole Shelton, "An Fmri Investigation Of The Impact Of Interracial Contact On Executive Function," *Nature Neuroscience*, November 2003, Volume 6 No 11, pp1323 – 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*, April 4, 2003.
- Stinebrickner, Todd and Ralph Stinebrickner, "Peer Effects Among Students From Disadvantaged Backgrounds," mimeo, 2001.
- Tuma, N.B. and Maureen T. Hallinan, "The Effects of Sex, Race and Achievement on Schoolchildren's Friendships," *Social Forces*, 1979.
- Waller, W., *The Family: A Dynamic Interpretation*. New York: Dryden, 1938.
- Winston, Gordon C., "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education," *Journal of Economic Perspectives*, vol. 13, no. 1, Winter 1999, pp. 13-36
- Zimmerman, David J., "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, vol. 85, no. 1, February 2003, pp. 9-23.

### Table 1: 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 a 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. All of 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	Freq.	Fraction Of Class
2003	491	0.47
2004	416	0.39
2005	236	0.23
2006	107	0.11
Total	1,250	

### Table 2: Summary Statistics for Primary Users

Primary users are those who use the NetBlitz system for email and have agreed to participate. All of the analysis that follows considers emails sent between primary users and the set of all students. N=1,250

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

### Table 3: 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.

Variable	Obs	Mean	Std. Dev.	Min	Max
Talks (0-1) (ie 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 person 1 and 2	4,225,623	0.253	9.831	0	3745
Volume sent conditional on volume $\geq$ 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 thousand of feet	4,225,623	1.421	0.839	0.060	3.273
Same major	4,225,623	0.069	0.254	0	1

**Table 4**  
**Mean Volumes and Probability 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 (total number of messages sent back and forth ) and the probability of talking (i.e. sending at least one message in each direction).

		<b>Same Dorm?</b>		<b>Ratio (Same Dorm to Not)</b>	
		No	Yes		
<b>Person 2 Black?</b>	<b>No</b>	Mean (Volume)	0.708	2.947	4.162
		Mean (Talks?)	0.017	0.054	3.176
		N	643,294	28,874	
	<b>Yes</b>	Mean (Volume)	0.212	0.619	2.920
		Mean (Talks?)	0.008	0.021	2.625
		N	48,182	2,220	

**Table 5**  
**Effects of Having A Black Roommate or**  
**Being Assigned to a Freshman Hall That is More than 6.5 Percent Black**

We ask whether there is any effect on interracial interaction from being assigned a black freshman roommate or a freshman hallway that has more than the median percent black. We do this by stratifying on own race and on having any black roommate freshman year (> median hallway % black). We calculate the percent of each student's volumes that is exchanged with black students and the probability of talking (percent of correspondents) that are black.

Potentially white students may see a direct effect on interracial contact from communicating with their black hallmates, and there is an potentially an indirect effect if your hallmates introduce you to additional black students. Using the regression tables that follow, we show that there is a strong direct effect and no indirect effect.

**Mean Volumes and Probability Of Talking**  
**By Own Race and Having a Black Roommate**

		<b>Has Black Roommate</b>	
		No	Yes
<b>Student Is Black</b>			
No	% volume to black students	0.044	0.049
	% of correspondents black	0.038	0.047
	N	1,097	92
Yes	% volume to black students	0.394	0.391
	% of correspondents black	0.452	0.386
	N	52	8

**Mean Volumes and Probability Of Talking**  
**By Own Race and Whether Freshman Hallway Has**  
**More than the Median Percent Black**

		<b>Hallway &gt; Median</b> <b>Percent Black?</b>	
		No	Yes
<b>Student Is Black</b>			
No	% volume to blacks students	0.042	0.047
	% of correspondents black	0.036	0.042
	N	688	498
Yes	% volume to blacks students	0.434	0.362
	% of correspondents black	0.515	0.388
	N	26	34

**Table 6**  
**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 and regress this fraction on the percent black on a student's freshman hallway, dorm, and a dummy for having a black roommate. We control for the student's characteristics including race, SAT scores, gender, fraternity membership, financial aid status, athlete status, type of high school attended.

	(1)	(2)	(3)
	Fraction	Fraction	Fraction
	Volume	Volume	Volume
	With Black	With Black	With Black
	Students	Students	Students
	(person 1 white)	(person 1 white)	(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 (0.032)*	0.083 (0.032)*	0.083 (0.032)*
Observations	900	900	900
R-squared	0.021	0.015	0.015

Standard errors in parentheses

\* significant at 5%; \*\* significant at 1%

**Table 7: 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 person 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  $e^{\text{coefficient}}$  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 as in Table 4.

	(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 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)**
Person2 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)**
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		-0.566	-0.731	-1.071	-0.547	-0.002

(Person 2)	(0.105)**	(0.207)**	(0.129)**	(0.162)**	(0.000)**
	[0.568]	[0.481]	[0.343]	[0.579]	
Both Are Athletes	1.297	1.296	1.553	1.019	0.010
	(0.142)**	(0.576)*	(0.343)**	(0.355)**	(0.001)**
	[3.658]	[3.655]	[4.726]	[2.770]	
Greek Member	-0.682	-0.042	-0.179	-0.319	-0.003
(Person 1)	(0.097)**	(0.218)	(0.217)	(0.247)	(0.000)**
	[0.506]	[0.959]	[0.836]	[0.727]	
Greek Member	-0.207	-0.514	-0.308	-0.107	-0.001
(Person 2)	(0.097)*	(0.190)**	(0.163)	(0.250)	(0.000)**
	[0.813]	[0.598]	[0.735]	[0.899]	
Both Are In Greek	1.382	1.859	1.008	0.842	0.010
Organizations	(0.131)**	(0.279)**	(0.254)**	(0.322)**	(0.001)**
	[3.983]	[6.417]	[2.740]	[2.321]	
Same Major	0.239	0.093	0.097	0.175	0.002
	(0.091)**	(0.166)	(0.194)	(0.202)	(0.000)**
	[1.270]	[1.097]	[1.102]	[1.191]	
Absolute Difference in	-0.003	-0.001	-0.001	0.000	0.000
SAT Scores	(0.000)**	(0.001)	(0.001)	(0.001)	(0.000)**
	[0.997]	[0.999]	[0.999]	[1.000]	
Person 1 is Male	-0.150	0.098	-0.148	0.166	0.000
	(0.069)*	(0.209)	(0.138)	(0.189)	(0.000)
	[0.861]	[1.103]	[0.862]	[1.181]	
Person 2 is Male	-0.333	-0.375	-0.393	-0.572	-0.001
	(0.049)**	(0.151)*	(0.109)**	(0.134)**	(0.000)**
	[0.717]	[0.687]	[0.675]	[0.564]	
Both are Male	-0.122	0.086	0.030	0.005	0.003
	(0.050)*	(0.154)	(0.109)	(0.137)	(0.000)**
	[0.885]	[1.090]	[1.030]	[1.005]	
Person 1 Went To NYC	0.565	-0.131	-0.647	-1.259	0.002
Exam School	(0.270)*	(0.178)	(0.250)**	(0.289)**	(0.001)*
	[1.759]	[0.877]	[0.524]	[0.284]	
Person 2 Went To NYC	-0.052	-0.045	-0.071	0.341	0.000
Exam School	(0.140)	(0.333)	(0.301)	(0.323)	(0.000)
	[0.949]	[0.956]	[0.931]	[1.406]	
Both 1 And 2 Went To	1.779	0.389	1.369	0.678	0.003
NYC Exam Schools	(0.643)**	(0.522)	(0.491)**	(0.359)	(0.002)
	[5.924]	[1.476]	[3.931]	[1.970]	
Person 1 Went	-0.247	-0.045	-0.725	0.026	0.000
To Fancy Prep School	(0.116)*	(0.216)	(0.239)**	(0.255)	(0.000)
	[0.781]	[0.956]	[0.484]	[1.026]	
Person 2 Went	0.173	-0.563	0.079	-0.008	0.000
To Fancy Prep School	(0.178)	(0.222)*	(0.190)	(0.246)	(0.000)
	[1.189]	[0.569]	[1.082]	[0.992]	
Both 1 And 2 Went	0.536	-0.320	-0.751	0.086	0.004
To Fancy Prep Schools	(0.284)	(0.734)	(0.731)	(0.651)	(0.001)**
	[1.709]	[0.726]	[0.472]	[1.090]	
Person 1 on Financial	-0.175	-1.101	0.148	-0.230	-0.001
Aid	(0.093)	(0.346)**	(0.213)	(0.264)	(0.000)**
	[0.839]	[0.333]	[1.160]	[0.795]	
Person 2 on Financial	-0.294	-0.583	0.241	-0.848	-0.001
Aid	(0.075)**	(0.408)	(0.221)	(0.260)**	(0.000)**
	[0.745]	[0.558]	[1.273]	[0.428]	
Both on Financial Aid	0.393	1.074	0.028	1.032	0.001
	(0.113)**	(0.469)*	(0.261)	(0.300)**	(0.000)**
	[1.481]	[2.927]	[1.028]	[2.807]	
Person 2 Is Black *	-0.291				-0.001
Person 1 Has Black	(0.312)				(0.001)
Roommate	[0.748]				

Person 2 Is Black* %		1.516	-3.501	1.624	4.791	0.002
Black In 1's Freshman		(1.749)	(3.765)	(4.063)	(2.681)	(0.004)
Dorm		[4.554]	[0.030]	[5.073]	[120.422]	
Person 1's Percent		-1.128				-0.003
Dorm Black		(0.752)				(0.002)
		[0.324]				
Person 1 Has		0.181				0.000
a Black Roommate		(0.144)				(0.000)
		[1.198]				
Constant	-2.242	-1.140	-2.129	-2.382	-1.666	
	(0.054)**	(0.103)**	(0.437)**	(0.258)**	(0.347)**	
Observations						
	3,021,484	2,923,120	192,279	560,120	245,165	2,923,120

Robust standard errors in parentheses. \* significant at 5%; \*\* significant at 1%. 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 20 feeder high schools to Dartmouth.

**Table 8: 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^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 Xs in 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 (0.160)** [3.108]	0.034 (0.005)**	1.857 (0.312)** [6.404]	0.971 (0.286)** [2.641]
Same Freshman Year Floor	0.838 (0.217)** [2.312]	0.011 (0.001)**	1.649 (0.375)** [5.202]	0.379 (0.375) [1.461]
Same Freshman Year Dorm	0.661 (0.225)** [1.937]	0.003 (0.001)**	0.289 (0.336) [1.335]	0.925 (0.408)* [2.522]
Same Freshman Year Cluster Of Dorms	0.237 (0.165) [1.267]	0.001 (0.000)**	0.308 (0.403) [1.361]	0.471 (0.316) [1.602]
Freshman Year Residential Distance In Thousands Of Feet	-0.088 (0.049) [0.916]	0.000 0.000	-0.070 (0.107) [0.932]	0.037 (0.072) [1.038]
Observations	2,923,120	2,923,120	230,920	1,094,550

Robust standard errors in parentheses

\* significant at 5%; \*\* significant at 1%

**Table 9: Number and Fraction Correspondents In versus Out of Group**

Here correspondents/ friends are defined as people with whom a student sends and receives at least one email. For each primary user in the sample, we calculate the fraction of their friends that are within a certain group (athletes, black students, football players). We also calculate the fraction of their email volume that is exchanged within the group.

	<b>Fraction Friends That Are In Reference Group (Black, Athlete, Football)</b>	<b>Fraction Emails to Reference Group Group (Black, Athlete, Football)</b>
<b>Primary User is.....</b>		
<b>Non-black</b>	0.044	0.038
<b>Black</b>	0.393	0.443
<b>Non-athlete</b>	0.183	0.167
<b>Athlete</b>	0.493	0.521
<b>Not football player</b>	0.007	0.007
<b>Football Player</b>	0.387	0.296

**Table 10**  
**Is Your Freshman Dorm Group A Good Proxy For Your Peer Group?**

Here we show the fraction of emails that are exchanged with freshman dormmates, by graduating class. The class of 2003 students are seniors at the time of data collection and the 2006 students are freshmen.

Person 1's Class	Fraction Overall Volume Exchanged with Freshman Dormmates	Fraction of Correspondents Who Are From Freshman Dorm
2006	0.194	0.210
2005	0.157	0.139
2004	0.095	0.072
2003	0.075	0.056

**Table 11**  
**Correlation Between Peer Average SATs and GPA Weighting By Actual Amount of Interaction Versus Unweighted Mean Of Dorm or Dorm-Race Group**

We calculate peers' mean SATs and mean GPA for each student, weighting by the number of emails sent and received with each peer campus wide. (Zero emails= zero weight placed on that peer.) We then correlate this interactions weighted measure of peer ability with measures more readily available in other studies, such as the simple mean among peers in your randomly assigned freshman dorm and the mean within your freshman dorm and racial group. Finally we calculate mean peer SAT (GPA) weighting by the predicted email volume, where we predict volume from a Poisson regressions of actual volume on all available Xs for both the student and her peers including race, geography, financial aid status, fraternity membership, athletic status etc. (ie the same regressions in Table 7 above.)

The message from this table is that in the process of constructing peer ability measures, conditioning on race and geography is significantly better than conditioning on geography alone. Conditioning on all available Xs is of course better still.

	SATs of Peers, Weighted By Actual Email Volume
Mean (SAT of Peers in Same Freshman Dorm and Class)	0.006
Mean (SAT of Peers w/ Same Freshman Dorm, Class and Race)	0.351
SAT of Peers Weighted By Predicted Email Volume	0.509
	GPA of Peers, Weighted By Actual Volume
Mean (GPA of Peers in Same Freshman Dorm and Class)	0.096
Mean (GPA of Peers w/ Same Freshman Dorm, Class and Race)	0.233
GPA of Peers Weighted By Predicted Email Volume	0.432

**Table 12**  
**Peer Effects in Fraternity Membership**

We estimate peer effects in the decision to join a Greek organization using three 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 emails weighted peer average with the average from the randomly assigned freshman dormmates. All regressions control for (but suppress coefficients for) race and gender dummies, financial aid status, SAT scores, feeder school status, and athletic status.

	(1)	(2)	(3)
	Joins Fraternity or Sorority (OLS)	Joins Fraternity or Sorority (OLS)	Joins Fraternity or Sorority (2SLS) instrument= average from freshman dormmates
Mean (Fraternity Status) Among Freshman Dormmates	0.494 (0.056)**		
Mean (Fraternity Status) Among Peers Weighted By Email Volume		1.035 (0.036)**	1.437 (0.137)**
Race, Gender and Demographic Controls?	yes	yes	yes
Observations	1250	1250	1250
R-squared	0.093	0.413	0.354

Standard errors in parentheses

\* significant at 5%; \*\* significant at 1%

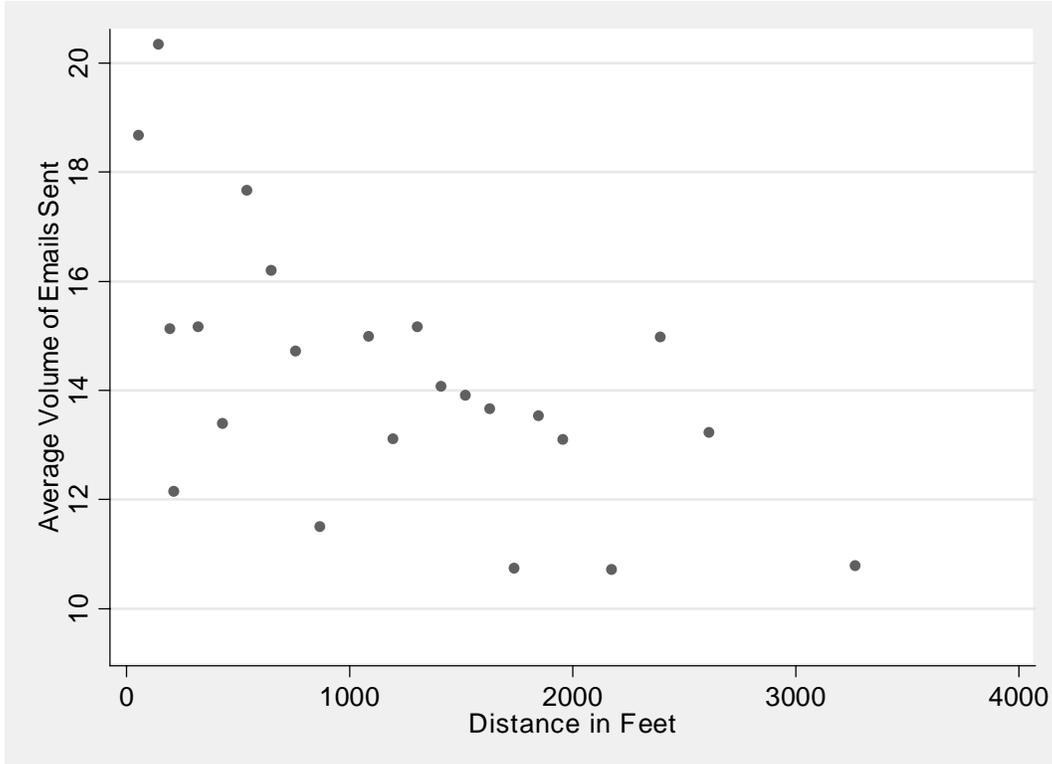
## Appendix 1: Comparison of Primary and Secondary Users

The Primary users are those that 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 Alfa users. The set of Alfa 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

### Figure 1 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.



**Figure 2**  
**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, we get a peer SAT distribution that matches the Actual line almost exactly.

