HOW FRIENDSHIPS FORM

David Marmaros

Google.com

Bruce Sacerdote[*]

Dartmouth College and NBER

bruce.sacerdote@dartmouth.edu

September 15, 2003

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 dataset. The data are from students and recent graduates of Dartmouth College. Our main finding is that geographic proximity and race are far more important determinants of friendship than are common interests or common majors. The effects of race are quite large; for example two randomly chosen black students are seven times more likely to interact than are a black student and a white student. Nonetheless, there still remain substantial amounts of interracial interaction, in part due to the powerful effects of distance coupled with randomized freshman housing. Women are more likely to interact with other women. But conditional on there being *any* communication between woman A and man B, the volume of communication between A and B is large. The results show that even short run residential mixing among people of different backgrounds promotes long run social interaction among those people.

I.   Introduction

It is often asserted 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 Case and Katz [1991], Evans Oates and Schwab [1992], Hoxby [2000a] and Sacerdote and Marmaros [2002]). Families intentionally self select into certain neighborhoods and students into certain schools because of the perceived peer effects (Hoxby [2000b], Winston [1999] and many others). However, much 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.[1]

One reason for the lack of studies on friendships 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].[2] We solve that particular problem by measuring the level of friendship, or at least 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, cultural similarity, 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 explained in Richards [2002], the universities' objectives may be partially blunted if the white

---

[1] As we discuss in the next section, there is a rich literature in psychology on the determinants of friendships. But our paper differs substantially from the existing work in that we are proposing a detailed and continuous measure of the level of social interaction and we are investigating the tradeoffs between geographic proximity, cultural and racial similarity, and common interests.

and non-white groups on campus spend very little time interacting. We show that campus segregation remains quite severe in 2003. But, forced racial mixing of freshman housing does a great deal to increase social interaction among individuals of different races.

In the recent Supreme Court cases addressing affirmative action, eight universities (Dartmouth, Harvard, Yale, Brown, University of Chicago, Duke, University of Pennsylvania and Yale) 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, Epstein [2002]) demonstrate that cross racial learning takes place and is valued by students and the labor market.[3]

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] on Bryn Mawr.

We add to the existing literature on friendship or social interactions in several ways. First, we have a much 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, and intellectual interests. We explore how the importance of these factors varies within versus across race and within versus across gender relationships. Finally, by looking at the same students over time, we explore how social interactions change following the students' departure from campus after graduation.

---

[2] 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.
[3] Tragically, no Sacerdote or Glaeser, Scheinkman Sacerdote papers on peer effects were cited.

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], and Kremer and Levy [2000] use college or university roommates to examine peer effects on both academic and social (particularly drinking) outcomes.

Like us, several authors including Festinger at 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 faculty 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 Ranio [1966] and Tuma and Hallinan [1978] imply that similarity and status are two important factors. Interestingly, the theorists in the literature, such as Waller [1938] and Blau [1964], model friendship much the way economists would; offers of friendship are made and accepted or rejected based on the costs and benefits of the relationship.

Modeling the friendship process

In conducting our analysis we have in mind a certain model of how friendships form and blossom. The idea is that every potential social interaction has associated costs and benefits. The benefits are both a flow of information and ideas and the utility from sharing a common experience and conversation with another human being. The costs are basically the time it takes to have the 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.[4]

Distance presents itself as a big cost, particularly when the value of the social interaction is unknown and especially if the person with whom the interaction might take place is unknown. Common background, interests and race between two people could raise or lower the benefits of a given social interaction. For example, a white senior from Newton, MA may have little in common with a black freshman from Chicago. This might increase the benefits of the interaction to both since the 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.

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

---

[4] Consider a hypothetical male student X who lives in the dorms farthest from the center of campus. In theory, X could walk .7 miles to the other 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 woman who would be open to a dinner invitation. But making this trip with no additional information, would be a very costly and probably embarrassing thing for X to do, particularly if the probability of success is low.

(# of emails sent from person 1 to person 2) = α + **B1***(dummies for person1's race, varsity athlete status, gender, Greek status) + **B2***(dummies for person 2's race, varsity athlete status, gender, Greek status) + β3*(dummy for same graduating class) + β4*(dummy for same freshman dorm) + **B5***(interactions of race and same freshman dorm) + **B6*** (interactions of female with same dorm dummy, race dummies)

We include person A sending to person B and B sending to A as two separate data points. However, our results are robust to combining the two volumes. Person A sending email to person B is not perfectly correlated with the reverse. Conditional on person A sending one or more emails, person B is about 78 percent likely to send one or more emails in the other direction.

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). 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.[5]

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 dropped all emails that were sent to more than one person. We assumed that such emails might be sent among friends, but they also are sent to working groups or perhaps to 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.[6]

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.

Table 1 shows a tabulation of the primary users by graduating class. Nearly half of the graduating class of 2003 (506 out of 1099) and roughly 40 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.

---

[5] In other words, we capture not just volumes sent during that particular session, but any information in the student's folders.
[6] Eight percent of NetBlitz users opted out of the study.

Table 2 shows some summary statistics on the primary users in the data set. On average the group is 48 percent male, 70 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 high school class rank (weight=1/3). On average the primary users sent 651 messages during the sample period with a standard deviation of 711 messages.

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 114 freshman are primary users and the freshman are prohibited from joining. Black students are underrepresented among the primary users; 5.4 percent of primary users are black versus 8.0 percent of the secondary users.

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 people select into our sample by choosing to use NetBlitz.

The analysis in this paper examines the volume of emails sent from each primary user to all other students. 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 1300+ primary users with 4000+ students which results in 5.3 million pairs of students. Any of these 5.3 million possible social connections might be active (have email traffic), though in fact about 75,000 of the connections are active during the sample period.

Table 3 shows summary statistics at the pair level. In the first row, we see that in 1.4 percent of the pairs, person 1 has sent one or more emails to person 2.[7] Conditional on any email being sent, 11.5 messages are sent on average, though the standard deviation is 52 messages and the range is enormous. Roughly 25 percent of the pairs consist of two members from the same graduating class. 1.2 percent of pairs are from the same class and the same freshman hallway. 4.1 percent of pairs are from the same class and same freshman dorm.

The pair's 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.

Figure 1 shows a histogram of the volume (number of messages) sent by primary users. About 30 percent of active pairs involve the sender sending a single message and 15 percent have a total volume of two messages sent. The distribution has a very long right hand tail with a maximum of 3087 messages (not shown on the histogram).

III.   Results

Our primary question is the relative importance of distance, race, and similar interests in determining who interacts with whom, and how much. We address this in Tables 4A, 4B, and 4C. Our measure of geographic closeness is an indicator variable for whether or not the two individuals in the pair were in the same freshman year dorm. Since freshman dorms are randomly assigned, this dummy variable is exogenous.

In Table 4A the dependent variable is a dummy for whether or not person 1 sends 1 or more emails to person 2. Recall that the mean of the dependent variable is .014, meaning that

---

[7] We define an indicator variable called "talks" which equals 1 if person 1 has sent 1 or more emails to person 2.

1.4 percent of the pairings are active. We run the regression separately by the race of the sender. The right hand side variables are dummies for the race of the recipient, dummies for sharing the same class year or same major, dummies for athletic status of person 1 and person 2, and dummies for being in a fraternity or sorority. The omitted category for person 2's race is always white. We interact the race dummies with a dummy for the same freshman dorm.

We show OLS results, though similar results obtain when we run probit regressions. Standard errors are corrected for clustering at the person 1 level. If person A and person B are both primary users, then we include both A sending to B and B sending to A as separate data points. Our results are robust to dropping one of the two pairs or combining the two pairs.

Column (1) is for observations in which the sender is white. The coefficient on "person 2 is black" is -.006, which means that white students are .6 percent less likely to send an email to a randomly chosen black student relative to a randomly chosen white student. This is a fairly large effect when measured against the mean of 1.4 percent of pairs being active. However, the effect of person 2's race is small relative to other coefficients in column (1). For example, the effect of same freshman dorm for white recipients is .058. Two white students are four times more likely to talk if they shared the same freshman dorm and class rather than just the same class. They are more than twice as likely to talk if they are the same class year versus not. This latter fact is unsurprising given that the seniors (the 2003s) have had four years of shared experiences together whereas the freshmen (the 2006s) have only been on campus for one year.

The interaction of black recipient and same freshman dorm has a coefficient of -.036, which means that 60 percent of the same freshman dorm effect is wiped out when we consider white students sending email to black students rather than whites to whites. However, even with this attenuation, there is still a same freshman dorm effect of .022 for the likelihood of white students emailing black students. The positive same dorm effect for white-black interactions is 3.7 times the size of the negative level effect for white-black interactions. The same dorm effect

for white-asian and white-hispanic interactions is nearly as large as the same dorm effect for white-white interactions.

Both people being in greek organizations nearly doubles the probability of the two people talking, relative to only person being a greek member. To summarize, for white senders and black recipients, the positive effects of same freshman dorm on the probability of communication appear to be much larger than the negative level effects of the recipient being black. White-asian and white-hispanic interactions occur with only slightly lower probability than white-white interactions. The effects of sharing a class year or both being greek are each 3 times greater than the black effect.

In column (2), we ask whether these same results hold when we look at black senders. In column (2) we see that there is an enormous effect of the recipient's race when the sender is black. A black sender is 6.2 percentage points more likely to send an email if the recipient is a randomly chosen black person versus a randomly chosen white person. (For the raw means in a 2x2 matrix, see Appendix 3). The interaction effect of black and same dorm is massive at 14.3. A black sender has about a 1.4 percent chance of interacting with a randomly chosen student. If the recipient were black, from the same class, and from the same freshman dorm, there is a 24.3 percent chance of the students interacting. (The sum of the relevant effects is .062+.018+.020+.143 = .243.)

There is still a statistically significant same dorm effect of .018 for black students sending to white students. We interpret this as saying that while black students are very strongly inclined towards interacting with other black students, geography remains quite important. Housing a black freshman with a white freshman does not overwhelm the strong black-black effect. But such mixed housing does increase the probability that the black student emails the white student by 180 percent. (Here we are comparing the .018 coefficient in column (2) with a baseline probability of black-white interaction of .01 from Appendix 3, panel A.)

Columns (3) and (4) are for asian and hispanic students respectively. Being in the same freshman dorm increases the likelihood of asian-white interaction and hispanic-white interaction by 4.0 percent and 4.3 percent respectively. But the same dorm effect for asian-asian interactions is 8.8 percent (adding the two relevant coefficients) and the same dorm effect for hispanic-hispanic interactions is 11.9 percent.

Table 4B mirrors Table 4A, except that the <u>number</u> of emails (including the zeroes) is the dependent variable, rather than a dummy for any email at all. Table 4C uses the log(number of emails) as the dependent variable. We drop all the observations with 0 emails sent, and hence are measuring the intensity of social interaction conditional on there being any interaction at all. We see in Table 4C that the relative importance of distance, race, and similar interests changes somewhat when we consider the intensity of social interactions among people.

Column (1) is for white senders. Conditional on emailing a black student, white students send 34 percent fewer emails than they send to white students. Being in the same class year increases the volume of interactions by 39 percent. If both students are white, sharing a freshman dorm increases volume by 15 percent. Interestingly, the same dorm effect for white-black volume is .04 above the .15 effect for white-white volume, though this interaction term is not statistically different from 0. Varsity athletes tend to be low volume senders and low volume recipients. However, if both persons 1 and 2 are athletes, the volume rises roughly to the level for a pair of non-athletes.

The fraternity/sorority effect works much the same way. Greek members send 16 percent less volume to their correspondents and greek recipients receive 8.5 percent less volume. But the coefficient on the greek-greek interaction term is positive 20 percent. Being in the same major raises volume by 6 percent for white senders and 15 percent for black senders.

Is the Distance Effect from Being Close, or Being Really, Really Close?

In Table 5, we take a closer look at the freshman dorm effect, and break it down by same freshman room, hallway, dorm, and cluster of dorms. A cluster at Dartmouth refers to a group of 2-4 dorm buildings that are located near each other and may be attached by breezeways and common rooms. We also add another measure of distance between the freshman year rooms of person 1 and person 2, measured in thousands of feet. We examine whether the distance effects degrade over time, by running separate regressions for each of the classes of 2003 through 2006.

Column (1) uses the whole sample and regresses the dummy for "talks" on dummies for same freshman room, floor, dorm, and cluster. The pattern in the coefficients on the dummies is striking. The effect of geographic closeness appears to decline exponentially with distance. Being freshman year roommates increases the probability of interacting by 31 percent (on top of the other same dorm effects) whereas being freshman hallmates or dormmates increases the chance of interaction by 7.6 percent and 1.6 percent respectively. Linear distance (in thousands of feet) appears to have no additional effect, once we control for the same room, hall, dorm, and cluster indicators.

Column (2) considers the log(volume) of email sent, conditional upon sending any. Here the pattern in the coefficients is still striking but not as extreme as in column (1). Being freshman roommates increases email volume by 31 percent. Being in the same cluster, dorm, or floor each add another 12-18 percent increase in email volume.[8]

In columns (3)-(6) we ask how much the effects of freshman dorm geography degrade over time. The freshman dorm effects have remarkable staying power. For the class of 2006, the effect of sharing a room on sending an email is 55 percent. For the sophomores, juniors, and seniors, this effect is roughly 26 to 33 percent. In other words, the coefficient drops from freshman year to sophomore year but appears relatively stable throughout sophomore-senior year. We add one caveat at this point which is that we only have 114 primary users among the

freshman and so the differences in coefficients may be driven by selection into the group of freshman primary users.[9] The effect of same freshman hall also appears to drop after the first year and then remain stable. The effect of same freshman dorm is similar for all four years.

In Tables 6A and 6B we examine relationships between men and women and ask whether the freshman dorm effect differs within versus across gender. We also ask whether inter-race relations differ within versus across gender.

In Table 6A we limit the sample to white male senders and ask how much person 2's gender matters and how that effect interacts with "same dorm" and race. The level effect of person 2 being a woman is -.4 percent and the same dorm effect is lowered by 2.5 percent for men. Men are significantly less likely to send an email to a randomly chosen woman than a man and the dorm effect is less strong across versus within gender.[10] However, in column (3) we see that conditional on sending any emails, men send 12 percent more volume to those women with whom they are in contact.

Column (2) shows that white men are slightly less likely to send an email to a person of another race. However, interacting female with race dummies, we see that white men are more likely to contact black or asian women than white men are likely to contact black or asian men.

In Table 6B, we limit the sample to white female senders and find very similar results. Women are less likely to talk to men, but they email more often conditional upon talking at all. The level effects of the black, asian, hispanic and other race dummies are negative, but the interaction of the race dummies with male are all positive.

---

[8] The dummies are overlapping so that for example, if the same floor dummy equals 1, the same dorm and same cluster dummy also equal 1 by definition. Thus the total effect of same floor is the sum of all three dummies.
[9] Obviously we also ignored the possibility of cohort effects. Experience leads us to think that the cohort effects would be less important than the age effects, but we do not offer any proof of this.
[10] This latter finding could be partially caused by the fact that all roommate pairs are same gender and I need to fix this.

How Does the Pattern of Interaction Change After Joining A Fraternity/Sorority or After Graduating?

In Table 7 we examine those sophomores who join a fraternity or sorority two academic quarters before and two academic quarters after joining. The point is to ask whether joining a fraternity or sorority is associated with any marked change in social interactions. The dependent variable is a dummy for "talks." The coefficients on the dummies for the recipient's race change very little upon joining a fraternity or sorority. Interestingly the effects of "same class" and "same dorm" increase markedly after joining. One interpretation is as follows: We know from Sacerdote [2001] that people in the same freshman dorm are likely to join the same Greek house. The Greek house might cause a big increase in traffic between individuals in the same house, particularly in their first year of joining.

Table 8 performs a similar analysis, but looks at the seniors one term before and one term after graduation. The sample is limited to white students.[11] The effects of recipient's race, freshman dorms, and similar interests (major, athletics) are remarkably similar both before and after graduation. In other words, initial effects of geography do not appear to be limited to one's time at Dartmouth, but continue on after graduation.

What Determines the Academic Ability of My Friends?

In Tables 9A and 9B we ask how my characteristics are correlated with the ability of those people with whom I interact. Our three measures of peer ability are incoming Academic Index, incoming math SATs, and cumulative GPA as of June 2003.

We take the sample of active pairs, i.e. where at least 1 email was sent during the year. In Table 9A, we regress the recipient's ability on the sender's characteristics. In Table 9B we include sender and recipient characteristics as regressors. In column (2) of Table 9A we show

that black students interact with students with lower math SAT scores. The effect is -36 SAT points which is about -.60 standard deviations. This effect increases to a full standard deviation (62 SAT points) when we weight by the volume of emails sent [column (4)].

Of course, the results in Table 9A could just be reflecting the fact that black students are more likely to interact with other black students (and athletes with other athletes). We address this issue in Table 9B by including dummies for the recipient's characteristics including race. Thus we can ask the question, do I interact with high (or low) ability students conditional on the race and athletic status of my correspondents? Controlling for recipient race and athletic status, there is still a significant though much smaller connection between own race and peers' academic ability. On average, black students interact with other students who have math SATs that are 6 points lower. Weighting for volume, as in column(2), increases this effect to -16 points.

The effects of being a varsity football player on the ability of own peer group are on the same order of magnitude as the effects of being black on peer group ability. In 9A column (2) we find that varsity athletes email students who have math SATs that are 7 points lower than non-athletes. Being a varsity football player further depresses peer SATs by 19 points. But in Table 9B, column (2) we see that controlling for the recipient's characteristics reduces the effect on peer SATs to a total of -3.4-2.4= -7 points.

Our bottom line is that while there is some negative effect of being black or being an athlete on peer academic ability, much of this negative effect can be explained simply by the fact that athletes (or minority students) have a predisposition to interact with each other. Athletes (or minority students) also interact with large numbers of non-athletes (non-minorities). And these out of group interactions are with peers of only slightly lower academic scores.

Appendix 3 contains some summary statistics on the likelihood that a black and non-black student interact. It is true that 7.2 percent of randomly chosen pairs black students interact,

---

[11] We limit the sample to whites to avoid having to fully interact dummies for the sender's race with every other

whereas only 1 percent of black – non-black pairs interact. However, given the large number of white, asian, and hispanic students in the student body, the majority of a black student's interactions are with someone of a different race.

For an average black student, 47 percent of his email messages are sent to other black students.[12] Thus black correspondents are overrepresented by a factor of 6. However, this still leaves 53 percent of his emails being sent to white, hispanic, asian and other race students. From the perspective of a non-black sender, 2.5 percent of her emails are sent to black students, leaving black students underrepresented relative to their 8 percent of the population. In other words, there is a very strong within race bias, but there is by no means complete segregation. And geographic proximity is a powerful force in stimulating cross-race interaction.

IV.   Conclusion

We find that geographic closeness, racial similarity, and common interests like academic majors, Greek organizations and varsity athletics all have large positive effects on the likelihood that two students interact. Using our results for white students sending email to black students, we find that the positive effects of being in the same freshman dorm (on interaction) are 3.7 times larger than the negative effect of the recipient being black. This indicates that thorough racial mixing of residences overcomes much of the existing tendencies towards racial segregation.

This result may have some broader implications outside of the university setting. It is possible that the effect of geographic closeness is just as important for neighbors in a city as for neighbors on a college campus. If true, then residential segregation in city must have a large negative effect on the likelihood that black and white residents meet each other and interact.

---

right hand side variable.

[12] These means are not shown in tables, but use the same data as all the results that are shown.

Policies which reduce residential segregation will inherently increase interracial interaction and knowledge, though of course not necessarily racial harmony.

Two white students have about a 1.4 percent chance of interacting whereas a white and a black student have a 1.0 percent chance of interacting. Conditional on emailing at all, white senders send 34 percent less volume to black correspondents than to white correspondents. Slightly more than half of each black student's correspondence is with non-black peers.

Black and white students do interact frequently, albeit at much lower rates than the within race interactions. This fact supports the argument of universities and policy makers who support affirmative action on the basis that it allows cross race learning and peer effects. Prior to this work, our greatest fear was that selective private universities were effectively running two separate campuses for black and non-black students, and that the two campuses had little interaction. The data show that the level of segregation is high, but not nearly that dire.

Do black students get "gypped" in college by interacting with peers of lower than average academic ability? We find that in the means, black students and athletes interact with peers who have significantly lower test scores and GPAs. However, the bulk of this effect is caused by the tendency for within group interactions to be more likely than out of group interactions. Focusing on the out of group interactions, black students and athletes email other students with only slightly lower SATs and GPAs.

Overall we conclude that the aggressive pro-diversity policies of universities have met at least one of their goals; white students do interact frequently with minority students. Much of the interracial interaction can be attributed to the very powerful effects of geographic closeness, which we show is a key determinant of who talks to whom.

References

Abu-Ghazzeh, 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.

Blau, P.M., *Exchange and Power in Social Life*, New York: Wiley, 1964.

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.

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.

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.

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.

Hallinan, Maureen T., "The Process of Friendship Formation," *Social Networks*, vol 1., 1978/9, pp. 193-210.

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, <u>A Thin Line Between Red and Green: Emerging Patterns in Group Preference an Exploration at Bryn Mawr College</u>, 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.

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.

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. Our data set includes nearly half of the classes of 2003 and 2004. All of the analysis that follows considers emails sent from the primary users to any other students on campus. The set of primary users is large enough that even the secondary users appear at least once in the data set, by virtue of sending or receiving email from a primary user.

Sender's class	Freq.	Fraction Of Class
2003	506	.46
2004	437	.39
2005	261	.22
2006	114	.11
Total	1,318	100

Table 2: Summary Statistics for Primary Users

Primary users are those who use the NetBlitz system for email and have agreed to participate. Our data set includes nearly half of the classes of 2003 and 2004. All of the analysis that follows considers emails sent from the primary users to any other student.

Variable	Obs	Mean	Std. Dev.	Min	Max
Male	1318	0.48	0.50	0	1
Member of fraternity/sorority	1318	0.49	0.50	0	1
White (0-1)	1206	0.70	0.46	0	1
Black	1206	0.05	0.22	0	1
Asian	1206	0.15	0.36	0	1
Hispanic	1206	0.07	0.25	0	1
Academic Index (from admissions)	1293	215.22	13.14	159.67	240.00
Cumulative GPA (as of 7/03)	1318	3.38	0.35	1.78	3.99
Math SAT Score	1304	715.46	61.78	500.00	800.00
Total Volume of Messages Sent	1318	651.32	710.56	50.00	7831.00

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) from each primary user to every other student on campus. In the labels below, person 1 is the primary user and person 2 is the potential recipient of email from that primary user.

Variable	Obs	Mean	Std. Dev.	Min	Max
Talks (0-1) (ie person 1 sends 1+ emails)	5,348,776	0.014	0.118	0	1
Volume sent from person 1 to 2	5,348,678	0.163	6.333	0	3087
Volume sent conditional on volume>=1	75,447	11.544	52.078	1	3087
Persons 1 and 2 are members of same class year	5,348,678	0.249	0.432	0	1
Same freshman floor	5,348,776	0.012	0.108	0	1
Same freshman dorm building	5,282,607	0.041	0.199	0	1
Same freshman year cluster of buildings	5,348,776	0.121	0.326	0	1
Distance between freshman rooms in thousand of feet	5,281,202	1.417	0.838	0.060	3.273
Same major	5,348,678	0.068	0.252	0	1

Table 4A: Regression of "Talks" on Sender and Recipient Characteristics

The dependent variable is a dummy which equals 1 if person 1 sent one or more emails to person 2. Each column conditions on the race of person 1, i.e. the sender. We show OLS coefficients, though partial derivatives from probits are nearly identical (results not reported). Standard errors use clustering at the sender level.

	(1) Sent at least 1 email	(2) Sent at least 1 email	(3) Sent at least 1 email	(4) Sent at least 1 email
	White Senders	Black Senders	Asian Senders	Hispanic Senders
Black recipient	-0.006 (0.000)**	0.062 (0.007)**	0.000 (0.001)	0.005 (0.002)*
Asian recipient	-0.004 (0.000)**	0.001 (0.001)	0.024 (0.002)**	-0.000 (0.001)
Hispanic Recipient	-0.002 (0.000)**	0.010 (0.002)**	0.001 (0.001)	0.011 (0.002)**
Other non-white Recipient	-0.007 (0.000)**	0.003 (0.002)*	-0.003 (0.001)**	-0.003 (0.001)*
Same class year	0.019 (0.001)**	0.018 (0.003)**	0.020 (0.002)**	0.020 (0.002)**
Same freshman Dorm	0.058 (0.003)**	0.020 (0.007)**	0.040 (0.005)**	0.043 (0.007)**
Black recipient* same freshman dorm	-0.036 (0.005)**	0.143 (0.030)**	0.006 (0.012)	0.019 (0.025)
Asian recipient* same freshman dorm	-0.016 (0.004)**	0.023 (0.017)	0.048 (0.012)**	-0.001 (0.015)
Hispanic recipient* same freshman dorm	-0.006 (0.006)	0.054 (0.026)*	0.015 (0.013)	0.076 (0.022)**
Other non-white recipient* same freshman dorm	-0.028 (0.006)**	0.053 (0.025)*	0.014 (0.016)	0.050 (0.028)
Varsity athlete (sender)	-0.006 (0.001)**	-0.004 (0.002)	0.001 (0.005)	-0.004 (0.003)
Varsity athlete (recipient)	-0.006 (0.000)**	-0.005 (0.001)**	-0.007 (0.000)**	-0.004 (0.001)**
Both are athletes	0.016 (0.001)**	0.009 (0.003)*	0.012 (0.003)**	0.011 (0.003)**
Greek member (sender)	-0.003 (0.001)**	0.004 (0.003)	0.001 (0.001)	-0.001 (0.002)
Greek member (recipient)	-0.002 (0.000)**	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)
Both are in Greek orgs	0.018 (0.001)**	0.014 (0.003)**	0.012 (0.001)**	0.019 (0.003)**

Same major	0.008	0.003	0.006	0.003
	(0.001)**	(0.002)	(0.002)**	(0.002)
Constant	0.011	0.002	0.004	0.007
	(0.000)**	(0.001)	(0.001)**	(0.001)**
Observations	3106164	236096	667756	306199
R-squared	0.014	0.031	0.018	0.014

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 4B: Regression of Volume on Sender and Recipient Characteristics

The dependent variable is the number of emails person 1 sent to person 2 during September 2002-July 2003. Each column conditions on the race of person 1, i.e. the sender. Regressions are OLS and standard errors use clustering at the sender level.

	(1) Number of Emails Sent	(2) Number of Emails Sent	(3) Number of Emails Sent	(4) Number of Emails Sent
	White Senders	Black Senders	Asian Senders	Hispanic Senders
Black recipient	-0.144	0.750	-0.021	-0.003
	(0.009)**	(0.121)**	(0.022)	(0.039)
Asian recipient	-0.076	0.016	0.453	0.005
	(0.013)**	(0.018)	(0.067)**	(0.046)
Hispanic Recipient	-0.046	0.118	-0.000	0.230
	(0.015)**	(0.037)**	(0.021)	(0.072)**
Other non-white Recipient	-0.097	0.040	-0.042	-0.026
	(0.019)**	(0.031)	(0.022)	(0.045)
Same class year	0.339	0.243	0.387	0.317
	(0.020)**	(0.033)**	(0.038)**	(0.047)**
Same freshman Dorm	1.023	0.670	1.192	1.691
	(0.122)**	(0.173)**	(0.261)**	(0.657)*
Varsity athlete (sender)	-0.109	-0.080	-0.068	-0.115
	(0.014)**	(0.025)**	(0.045)	(0.030)**
Varsity athlete (recipient)	-0.112	-0.085	-0.118	-0.035
	(0.014)**	(0.020)**	(0.013)**	(0.032)
Both are athletes	0.249	0.127	0.184	0.134
	(0.021)**	(0.040)**	(0.061)**	(0.061)*
Greek member (sender)	-0.062	-0.022	-0.041	-0.033
	(0.011)**	(0.017)	(0.030)	(0.037)
Greek member (recipient)	-0.034	-0.003	-0.062	-0.027
	(0.017)*	(0.018)	(0.029)*	(0.041)
Both are in Greek orgs	0.231	0.278	0.158	0.233
	(0.027)**	(0.043)**	(0.038)**	(0.064)**
Same major	0.071	0.011	0.057	-0.018
	(0.030)*	(0.028)	(0.054)	(0.049)
Constant	0.139	-0.014	0.051	0.067
	(0.011)**	(0.015)	(0.022)*	(0.025)*
Observations	3106164	236096	667756	306199
R-squared	0.001	0.006	0.002	0.002

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 4C: Regression of Log(volume) on Sender and Recipient Characteristics

The dependent variable is the log(# emails) person 1 sent to person 2 during September 2002-July 2003. Pairs with 0 emails sent are dropped. Each column conditions on the race of person 1, i.e. the sender. Regressions are OLS and standard errors use clustering at the sender level.

	(1) Log(volume sent)	(2) Log(volume sent)	(3) Log(volume sent)	(4) Log(volume sent)
	White Senders	Black Senders	Asian Senders	Hispanic Senders
Black recipient	-0.346 (0.026)**	0.498 (0.092)**	-0.157 (0.064)*	-0.095 (0.076)
Asian recipient	-0.136 (0.023)**	-0.027 (0.097)	0.281 (0.042)**	-0.095 (0.079)
Hispanic Recipient	-0.048 (0.027)	0.243 (0.106)*	-0.045 (0.055)	0.204 (0.082)*
Other non-white Recipient	-0.195 (0.046)**	-0.012 (0.130)	-0.067 (0.095)	-0.085 (0.113)
Same class year	0.387 (0.019)**	0.261 (0.070)**	0.415 (0.038)**	0.341 (0.049)**
Same freshman Dorm	0.153 (0.036)**	0.213 (0.147)	0.382 (0.098)**	0.370 (0.177)*
Black recipient* same freshman dorm	0.042 (0.148)	-0.066 (0.239)	0.352 (0.317)	-0.134 (0.265)
Asian recipient* same freshman dorm	0.236 (0.096)*	0.372 (0.400)	-0.391 (0.131)**	0.040 (0.350)
Hispanic recipient* same freshman dorm	0.132 (0.131)	0.519 (0.341)	-0.143 (0.256)	-0.692 (0.259)**
Other non-white recipient* same freshman dorm	0.402 (0.189)*	0.672 (0.380)	0.138 (0.259)	-0.034 (0.422)
Varsity athlete (sender)	-0.143 (0.031)**	-0.270 (0.098)**	-0.198 (0.106)	-0.367 (0.092)**
Varsity athlete (recipient)	-0.172 (0.021)**	-0.229 (0.063)**	-0.149 (0.037)**	-0.093 (0.052)
Both are athletes	0.313 (0.035)**	0.384 (0.141)**	0.191 (0.094)*	0.296 (0.095)**
Greek member (sender)	-0.160 (0.036)**	-0.253 (0.118)*	-0.123 (0.059)*	-0.098 (0.076)
Greek member (recipient)	-0.085 (0.024)**	-0.045 (0.059)	-0.063 (0.055)	-0.108 (0.072)
Both are in Greek orgs	0.199 (0.037)**	0.506 (0.111)**	0.061 (0.071)	0.182 (0.095)

Same major	0.063	0.153	0.023	-0.032
	(0.030)*	(0.081)	(0.050)	(0.070)
Constant	1.064	0.735	0.972	1.031
	(0.023)**	(0.054)**	(0.048)**	(0.059)**
Observations	43869	3405	9253	4742
R-squared	0.038	0.080	0.049	0.038

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 5: The Effect of Distance on the Amount of Interaction
And How The Effects of Distance Degrade with Time

The dependent variables are a dummy for "has sent 1 or more emails" or the log(# emails) person 1 sent to person 2 during September 2002-July 2003.   Regressions are OLS and standard errors use clustering at the sender level. Columns (3)-(6) show the baseline regression conditioning on the class of the sender. The objective is to examine the degree to which the importance of freshman dorm assignment decreases over time.

	(1) Sent at least 1 email (whole sample)	(2) Log(volume sent) (whole sample)	(3) Sent at least 1 email Class of 2006	(4) Sent at least 1 email Class of 2005	(5) Sent at least 1 email Class of 2004	(6) Sent at least 1 email Class of 2003
Same freshman year room	0.311 (0.014)**	0.314 (0.073)**	0.551 (0.045)**	0.336 (0.032)**	0.259 (0.024)**	0.310 (0.022)**
Same freshman year floor	0.076 (0.004)**	0.114 (0.055)*	0.146 (0.014)**	0.086 (0.007)**	0.064 (0.006)**	0.054 (0.007)**
Same freshman year dorm	0.016 (0.002)**	0.168 (0.046)**	0.020 (0.005)**	0.019 (0.004)**	0.013 (0.003)**	0.014 (0.003)**
Same freshman year cluster of dorms	0.022 (0.001)**	0.177 (0.029)**	0.006 (0.003)	0.008 (0.002)**	0.002 (0.002)	0.007 (0.002)**
Freshman year residential distance in thousands of feet	0.000 (0.000)	-0.002 (0.008)	-0.002 (0.001)*	0.000 (0.001)	-0.004 (0.001)**	0.000 (0.001)
Constant	0.013 (0.000)**	1.116 (0.016)**	0.020 (0.002)**	0.022 (0.002)**	0.035 (0.002)**	0.034 (0.001)**
Observations	5281196	74762	113328	272484	426530	503577
R-squared	0.008	0.009	0.045	0.016	0.007	0.007

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 6A: Men Talking to Women, By Woman's Race

The dependent variables are a dummy for "has sent 1 or more emails" or the log(# emails) person 1 sent to person 2 during September 2002-July 2003. <u>The sample of senders is limited to white men</u>. We include a dummy for female recipient and interact this with indicators for the recipient's race. Regressions are OLS and standard errors are clustered at the sender level.

	(1) Sent at least 1 email	(2) Sent at least 1 email	(3) Log(volume sent)
Female recipient	-0.004	-0.004	0.122
	(0.000)**	(0.000)**	(0.028)**
Same freshman year dorm	0.079	0.081	0.526
	(0.003)**	(0.004)**	(0.057)**
Female* same Freshman Dorm	-0.025	-0.029	-0.330
	(0.004)**	(0.005)**	(0.103)**
Black female recipient		0.001	-0.136
		(0.001)*	(0.081)
Asian female recipient		0.002	-0.065
		(0.001)**	(0.060)
Hispanic female recipient		-0.001	-0.082
		(0.001)	(0.076)
Other non-white female Recipient		0.000	-0.018
		(0.001)	(0.122)
Black recipient		-0.007	-0.303
		(0.001)**	(0.050)**
Asian recipient		-0.005	-0.069
		(0.001)**	(0.038)
Hispanic recipient		-0.001	-0.054
		(0.001)	(0.044)
Other non-white recipient		-0.006	-0.122
		(0.001)**	(0.080)
Constant	0.014	0.015	1.085
	(0.000)**	(0.001)**	(0.023)**
Observations	2,538,557	1,523,529	19,361
R-squared	0.004	0.005	0.013

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 6B: Women Talking to Men, By Man's Race

The dependent variables are a dummy for "has sent 1 or more emails" or the log(# emails) person 1 sent to person 2 during September 2002-July 2003. The sample of senders is limited to white women. We include a dummy for male recipient and interact this with indicators for the recipient's race. Regressions are OLS and standard errors are clustered at the sender level.

	(1) Sent at least 1 email	(2) Sent at least 1 email	(3) Log(volume sent)
Male recipient	-0.007	-0.010	0.044
	(0.000)**	(0.001)**	(0.028)
Same freshman dorm	0.087	0.087	0.523
	(0.003)**	(0.004)**	(0.060)**
Male* same Freshman Dorm	-0.039	-0.035	-0.296
	(0.004)**	(0.005)**	(0.091)**
Black male recipient		0.010	0.032
		(0.001)**	(0.082)
Asian male recipient		0.004	0.027
		(0.001)**	(0.065)
Hispanic male recipient		0.004	0.089
		(0.001)**	(0.080)
Other male recipient		0.007	0.123
		(0.001)**	(0.147)
Black recipient		-0.014	-0.387
		(0.001)**	(0.046)**
Asian recipient		-0.009	-0.153
		(0.001)**	(0.037)**
Hispanic recipient		-0.005	-0.036
		(0.001)**	(0.045)
Other non-white recipient		-0.013	-0.258
		(0.001)**	(0.060)**
Constant	0.018	0.022	1.173
	(0.000)**	(0.001)**	(0.021)**
Observations	2743952	1582635	24508
R-squared	0.005	0.006	0.011

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 7: Email Behavior of Sophomores, Before and After Joining A Fraternity or Sorority

The dependent variable is a dummy which equals 1 if person 1 sent one or more emails to person 2. Standard errors use clustering at the sender level. Column (1) is for the two terms prior to the sender joining a fraternity or sorority and column (2) is for the two terms after joining.

	(1) Volume sent before joining Greek house	(2) Volume sent after joining Greek house
Black recipient	-0.020	-0.024
	(0.010)*	(0.022)
Asian recipient	-0.010	0.010
	(0.011)	(0.028)
Hispanic recipient	-0.004	0.004
	(0.014)	(0.024)
Other non-white recipient	-0.029	-0.051
	(0.011)**	(0.014)**
Same class year	0.035	0.144
	(0.008)**	(0.018)**
Same freshman dorm	0.172	0.526
	(0.049)**	(0.108)**
Varsity athlete (sender)	-0.017	-0.053
	(0.015)	(0.018)**
Varsity athlete (recipient)	-0.010	-0.044
	(0.014)	(0.018)*
Both are athletes	0.030	0.153
	(0.016)	(0.042)**
Greek member (sender)	0.000	0.000
	(0.000)	(0.000)
Greek member (recipient)	0.048	0.094
	(0.019)*	(0.018)**
Both are in Greek organizations	0.000	0.000
	(0.000)	(0.000)
Same major	0.012	0.038
	(0.011)	(0.014)**
Male recipient	-0.002	-0.028
	(0.007)	(0.009)**
Constant	0.016	0.044
	(0.009)	(0.012)**
Observations	486856	486856
R-squared	0.000	0.001

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 8: Email Behavior of Seniors, Before and After Graduation

The dependent variables are the volume and log(volume) for the term before and the term after graduation. Regressions are OLS with standard errors clustered at the sender level. The sample is limited to senders who are white men and women in the class of 2003.

	(1) Volume sent Senior Spring	(2) Volume sent after Graduation	(3) Log(volume sent) Senior Spring	(4) Log(volume sent) After Graduating
Black recipient	-0.052	-0.034	-0.292	-0.379
	(0.006)**	(0.003)**	(0.061)**	(0.083)**
Asian recipient	-0.036	-0.016	-0.176	-0.115
	(0.007)**	(0.004)**	(0.049)**	(0.046)*
Hispanic recipient	-0.031	-0.014	-0.048	-0.031
	(0.008)**	(0.005)*	(0.057)	(0.063)
Other non-white recipient	-0.050	-0.025	-0.031	0.068
	(0.007)**	(0.007)**	(0.117)	(0.177)
Same class year	0.126	0.091	0.437	0.390
	(0.013)**	(0.007)**	(0.035)**	(0.038)**
Same freshman dorm	0.204	0.230	0.152	0.096
	(0.064)**	(0.065)**	(0.084)	(0.066)
Varsity athlete (sender)	-0.044	-0.023	-0.197	-0.202
	(0.008)**	(0.005)**	(0.061)**	(0.051)**
Varsity athlete (recipient)	-0.052	-0.025	-0.193	-0.210
	(0.009)**	(0.004)**	(0.047)**	(0.043)**
Both are athletes	0.096	0.049	0.280	0.274
	(0.013)**	(0.006)**	(0.071)**	(0.062)**
Greek member (sender)	-0.022	-0.020	-0.100	-0.076
	(0.006)**	(0.005)**	(0.069)	(0.063)
Greek member (recipient)	-0.013	-0.018	-0.044	-0.018
	(0.014)	(0.007)*	(0.055)	(0.059)
Both are in Greek organizations	0.068	0.058	0.008	0.105
	(0.018)**	(0.009)**	(0.072)	(0.079)
Same major	0.071	0.032	0.088	-0.037
	(0.017)**	(0.010)**	(0.055)	(0.054)
Constant	0.057	0.032	0.956	0.925
	(0.007)**	(0.005)**	(0.052)**	(0.051)**
Observations	1129187	1129187	7572	5961
R-squared	0.001	0.001	0.046	0.039

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 9A: How Own Characteristics Correlate With Peer Ability

The dependent variables are three measures of recipients' academic ability. We take the set of pairings for which one or more emails were sent. We regress the recipient's academic index, math SAT or cumulative GPA on sender characteristics. The academic index is the sum of math SAT, verbal SAT and an index for high school class rank. Standard errors are clustered at the sender level.   Columns (4) and (5) are weighted by the number of messages sent.

	(1) Recipient's academic index	(2) Recipient's math SAT	(3) Recipient's cumulative GPA	(4) Recipient's math SAT (weighted by volume)	(5) Recipient's cumulative GPA (weighted by volume)
Black sender	-8.890	-36.729	-0.151	-61.704	-0.250
	(0.904)**	(3.627)**	(0.021)**	(6.952)**	(0.030)**
Asian sender	0.927	6.489	0.004	6.926	0.003
	(0.224)**	(0.949)**	(0.007)	(2.868)*	(0.021)
Hispanic sender	-2.800	-11.731	-0.072	-21.265	-0.107
	(0.474)**	(2.007)**	(0.012)**	(5.199)**	(0.034)**
Other non-white Sender	-5.535	-18.093	-0.128	-19.428	-0.078
	(1.333)**	(4.567)**	(0.032)**	(8.042)*	(0.056)
Varsity athlete (sender)	-2.486	-6.714	-0.043	-8.660	-0.054
	(0.281)**	(0.981)**	(0.007)**	(2.636)**	(0.020)**
Greek member (sender)	-0.691	-1.903	-0.037	-4.499	-0.074
	(0.214)**	(0.841)*	(0.006)**	(2.336)	(0.017)**
Football player (sender)	-7.452	-18.809	-0.181	-19.074	-0.082
	(0.974)**	(4.095)**	(0.025)**	(8.027)*	(0.071)
Constant	215.839	716.920	3.398	724.443	3.446
	(0.163)**	(0.633)**	(0.004)**	(1.980)**	(0.013)**
Observations	67365	67846	68704	67846	68704
R-squared	0.036	0.027	0.019	0.056	0.037

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 9B: How Own Characteristics Correlate With Peer Ability
Controlling for Recipient's Race, Athletic Status, Greek Status

The dependent variables are three measures of recipients' academic ability. We take the set of pairings for which one or more emails were sent. We regress the recipient's academic index, math SAT or cumulative GPA on sender characteristics. Standard errors are clustered at the sender level. Columns (2), (4) and (6) are weighted by the number of messages sent. In contrast to Table 6A, this table controls for the recipient's characteristics.

	(1) Recipient's math SAT	(2) Recipient's math SAT (weighted)	(3) Recipient's academic index	(4) Recipient's academic index (weighted)	(5) Recipient's cumulative GPA	(6) Recipient's cumulative GPA (weighted)
Black sender	-6.405	-16.210	-1.179	-4.077	-0.006	-0.039
	(1.165)**	(4.059)**	(0.313)**	(0.877)**	(0.014)	(0.025)
Asian sender	1.563	3.310	0.386	-0.591	0.000	-0.013
	(0.756)*	(2.862)	(0.174)*	(0.660)	(0.006)	(0.018)
Hispanic sender	-5.029	-11.096	-1.019	-1.964	-0.034	-0.054
	(1.277)**	(4.339)*	(0.250)**	(1.003)	(0.009)**	(0.030)
Other non-white Sender	-6.667	-9.469	-2.165	-1.653	-0.055	-0.019
	(2.640)*	(5.898)	(0.675)**	(1.306)	(0.019)**	(0.037)
Varsity athlete (sender)	-3.402	-5.305	-1.348	-1.288	-0.025	-0.020
	(0.746)**	(2.745)	(0.190)**	(0.665)	(0.005)**	(0.024)
Football player (sender)	-2.418	-14.036	-1.825	-1.202	-0.069	-0.041
	(3.222)	(8.782)	(0.833)*	(2.178)	(0.017)**	(0.059)
Football player (recipient)	-34.505	-30.620	-11.914	-13.639	-0.215	-0.168
	(2.213)**	(9.143)**	(0.540)**	(1.719)**	(0.015)**	(0.037)**
Greek member (sender)	-1.532	-4.505	-0.372	-0.383	-0.016	-0.034
	(0.596)*	(2.315)	(0.136)**	(0.394)	(0.004)**	(0.015)*
Black recipient	-97.595	-100.013	-25.485	-24.981	-0.493	-0.474
	(1.059)**	(3.759)**	(0.206)**	(0.800)**	(0.007)**	(0.025)**
Asian recipient	14.060	9.579	-0.722	-0.405	-0.077	-0.046
	(0.622)**	(2.790)**	(0.127)**	(0.494)	(0.004)**	(0.018)**
Hispanic Recipient	-54.825	-58.865	-14.564	-14.536	-0.328	-0.326
	(1.079)**	(4.565)**	(0.196)**	(0.964)**	(0.006)**	(0.026)**
Other non-white Recipient	-77.548	-63.121	-19.394	-17.700	-0.441	-0.394
	(2.441)**	(15.824)**	(0.578)**	(2.042)**	(0.013)**	(0.044)**
Varsity athlete (recipient)	-18.963	-14.787	-7.386	-6.815	-0.154	-0.165
	(0.633)**	(2.982)**	(0.175)**	(0.572)**	(0.004)**	(0.029)**
Greek member (recipient)	-0.975	1.155	-0.888	-1.230	-0.070	-0.105
	(0.519)	(2.139)	(0.106)**	(0.367)**	(0.003)**	(0.014)**
Constant	729.406	732.794	220.545	221.491	3.525	3.563
	(0.521)**	(2.021)**	(0.117)**	(0.368)**	(0.004)**	(0.011)**
Observations	62238	62238	61915	61915	63025	63025
R-squared	0.260	0.239	0.360	0.348	0.206	0.198

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Figure 1
Histogram of Volume of Emails Sent Conditional on Volume>=1

This shows the volume of emails sent from Person 1 to 2 during June 2002 through July 2003. The data are truncated at a volume of 200 emails, though the maximum is actually 1300+.