The Old Boy (and Girl) Network:
Social Network Formation on University Campuses

Adalbert Mayer and Steven L. Puller [1]

June 2007

Abstract

This paper documents the structure and composition of social networks on university campuses and investigates the processes that lead to their formation. We use a large dataset that identifies students in one another's social network on campus and link these data to university records on each student's demographic and school outcome characteristics. The campus networks exhibit common features of social networks, such as clusteredness. We document the factors that are the strongest predictors of whether two students are friends. Race is strongly related to social ties. In particular, blacks and Asians have disproportionately more same race friends than would arise from the random selection of friends, even after controlling for a variety of measures of socioeconomic background, ability, and college activities. We show that academic and social outcomes of students are associated with outcomes of their friends. Next, we develop a model of the formation of social networks that decomposes the formation of social links into effects based upon the exogenous school environment and effects of endogenous choice arising from preferences for certain characteristics in one's friends. We use student-level data from an actual social network to calibrate the model. We simulate the social network under alternative university policies aimed at reducing social segmentation. We find that changes in the school environment that affect the likelihood that two students interact have only a limited potential to reduce the segmentation of the social network.

JEL Classification: I20, J15, J62, Z13

[1] Department of Economics, Texas A&M University. Email: amayer@econmail.tamu.edu and puller@econmail.tamu.edu We are grateful to Larry Malota, Dustin Moscovitz, and Arnie Vedlitz for their time and assistance in acquiring our data. Ahmad Alwaked, Manuel Hernandez, Travis Miller, Lauren Rozanski, and Joseph Wood provided excellent research assistance. We thank two anonymous referees, Philip Babcock, Julian Betts, Tim Gronberg, Dan McFarland, Kimon Ioannides, John Moroney, Bruce Sacerdote, Adriaan Soetevent, Lori Taylor, Manuelita Ureta, and seminar participants at the 2006 SOLE meetings, the NBER Education Program meeting, Texas A&M, Rice/University of Houston, the Wegmans Conference at Rochester, University of Augsburg, Western Ontario, the Bush School, Ohio State, University of Texas, University of Montreal, UNC-Greensboro, and University of Calgary for their helpful suggestions.

I. Introduction

Universities are important venues for the formation of social networks. For many students, college life is the first experience outside the environment determined by their parents. It directly precedes entrance into the job market and the resulting social contacts are an important channel of information transmission[2]. Connections between business partners are formed using knowledge from prior social interaction[3]. Employers and employees frequently use social contacts to obtain information about each other, and this can have important impacts on labor markets[4].

Interaction between members of different social groups in college is sometimes viewed as a policy goal in itself. It is argued that interaction between students from different backgrounds and walks of life provides a better learning environment to prepare students for an increasingly diverse workforce and society (see Bok and Bowen, 1998).[5] Consequently, universities have made concerted efforts over the last several decades to create diverse campuses that bring together students of different races and from diverse socioeconomic backgrounds. In a recent ruling (Grutter v. Bollinger, 2003) the United States Supreme Court upheld the right of universities to use race as an admissions criterion to increases diversity. One of the arguments

---

[2] For surveys of the literature on social interactions, see Ioannides and Loury (2004), Manski (2000), and Soetevent (2004).

[3] For example Cohen et. al. (2007) investigate "connections between mutual fund managers and board corporate board members via shared education networks." They find that "portfolio mangers place larger bets on firms they are connected to through their network," and suggest that "social networks may be an important mechanism for information flow into asset prices."

[4] Montgomery (1991) reviews several studies about the importance of social connections in the labor market and concludes that: "While the frequency of alternative job-finding methods varies somewhat by sex and occupation, the following generalization seems fair: approximately 50% of all workers currently employed found their jobs through friends and relatives" Pellizarri (2004) documents this phenomenon for various countries. Social networks affect whether jobs are filled efficiently and influence the incentives for investment in education and work force participation. Calvo-Armengol and Jackson (2004) show that information transmission in social networks can account for differences in wages and labor market participation of different social groups. In another paper Calvo-Armengol and Jackson (2006) show that the structure and composition of social networks have implications for intergenerational mobility.

[5] Minorities who maintain racially segregated social connections may fail to build social capital and networks that impact educational outcomes, future labor market outcomes, and social mobility. Moreover, all ethnic groups that remain racially segregated forego the opportunity to be exposed to diversity of thought from students with different cultural, socioeconomic and political perspectives.

against the current practice is the disconnect between the admissions policies and the actual campus experience.[6]

Indeed, it is not known whether a diverse university population leads to diverse interaction among students. Anecdotal evidence suggests that students form cliques based upon race or social background – a casual walk through campus of a large state university or a visit to the dorm cafeteria will illustrate social segmentation. There are few large scale empirical studies that quantify this pattern. A notable exception is Marmaros and Sacerdote (2006), who use data on email communication between Dartmouth students and find that race and distance on campus are important determinants of social interaction.[7]

In this paper, we use a large new dataset from 10 public and private universities to describe social networks in college. We make two contributions to the literature. First, we use student-level data to document the structure of these networks and to measure segmentation of social ties by race, socioeconomic background, and ability. We show that academic and social outcomes of students are associated with outcomes of their friends. Second, we develop a model of network formation that yields a network with many of the commonly observed characteristics of social networks. We calibrate the model to our data and perform counterfactual experiments of university policies that promote student diversity.

Our data are from Facebook.com, a student social networking website for each university. Students use this website to share information and stay in contact with each other. One feature of the Facebook identifies friendships between students and we exploit this information to measure students' social connections on campus.

---

[6] In his dissenting opinion Justice Scalia states: "Still other suits may challenge the bona fides of the institution's expressed commitment to the educational benefits of diversity that immunize the discriminatory scheme in *Grutter*. (Tempting targets, one would suppose, will be those universities that talk the talk of multiculturalism and racial diversity in the courts but walk the walk of tribalism and racial segregation on their campuses…"

[7] For analyses of secondary school social interaction and connectedness, see Joyner and Kao (2000), Moody (2001), Quillian and Campbell (2003), Weinberg (2005), Fryer and Torelli (2006), and Babcock (2006).

We find that social networks differ substantially from the network that would arise from random selection of friends. The structure of these networks exhibits the classic characteristics of social networks (see Jackson 2006). Social networks on campus are cliquish and exhibit a cluster coefficient that cannot be explained by random formation of links. The distribution of number of social connections varies more than implied by random links and it is right skewed. Agents with many ties tend to be connected with other agents with many ties.

At all 10 universities, similar characteristics of two students make the formation of a friendship more likely. Despite the fact these schools are very different in size and type, we find similar overall patterns in social segmentation. Sharing the same major or the same political orientation makes the formation of a friendship more likely. Students are twice as likely to form friendships within their cohort, as with students from the cohort above or below them. Friendships are more likely to be formed within the same race. The share of Asians among the friends of an Asian student is at least twice as high as the share in the total student population; at three of the ten universities it is more than seven times as high. For African American students the difference is even more pronounced. The share of African American friends of African Americans is 4.5 times (at Rice) to 15 times (at Texas A&M and University of Texas at Austin) as high as their share in the overall population.

We link the social network data to student-level administrative data for one of the universities -- Texas A&M. This allows us to use additional information on parental education and income, student SAT scores and college GPA, and the student's high school, dorm, and activities, such as athletics and fraternity/sorority membership.

Using this rich dataset on Texas A&M students, we investigate the determinants of friendship formation. First, we explore in a reduced-form setting the demographic and socioeconomic factors that are good predictors of two students becoming friends. Second, we use the reduced-form results to develop a model of friendship formation that we calibrate to our data from Texas A&M.

We estimate a linear probability model of any two students being friends. Relative to the baseline rate that any two students chosen at random are friends, students living in the same dorm are 13 times more likely to be friends, two black students are 17 times more likely to be friends, two Asian students are 5 times more likely to be friends, and two Hispanic students are about twice as likely to be friends. Socioeconomic background and academic achievement affect the probability of a friendship formation to a smaller but statistically significant degree.

Even though observable characteristics such as race clearly play a role in friendship formation, they have very little explanatory power for the formation of a friendship *between two specific students*. However, common friends are a good predictor for the existence of a friendship between two students – students *i* and *j* are much more likely to be friends if each is friends with student *k*. Moreover, when we control for the number of common friends, the importance of other characteristics such as race changes. Therefore, it is not sufficient to use the results of the linear probability model to predict the effects of changes in the determinants of the network. A better model of the effects of changes in school environment must incorporate how common friends can affect the formation of a friendship. Endogenous effects through friends of friends may magnify effects of a change in the university environment.

We build a model of friendship formation. This allows us to simultaneously study the environment that determines contact, tastes that determine the formation of friendships conditional on contact, and the influence of interconnections through friends of friends. Borrowing from the random graph literature[8], we specify a framework of random interactions of individuals. Like Jackson and Rogers (forthcoming), we allow links to form with any other student ("random attachment") and links to form through friends of existing friends ("preferential

[8] Building on the random graph literature, a number of mechanical stochastic processes of network formation have been proposed. These models are able to explain various features of social networks. Contributions can be found in the computer science, physics and economics literature (See Newman (2003)).

attachment" or "search"). In addition, we model heterogeneity in the environment and preferences of agents.

Our model starts with a network in which no students are connected through friendship. Then, two individuals meet with a probability that is determined by their school environment (e.g. dorm assignment or cohort). Conditional upon meeting, the students choose whether or not to form a friendship based upon tastes for observable characteristics. Finally, students meet friends of their friends and again choose whether to form a friendship based upon preferences. We calibrate the model to yield a simulated network that resembles the observed social network.

We simulate modified versions of the model to generate counterfactuals. We assess the effectiveness of policies that try to decrease socialization *within* subgroups and increase socialization *across* subgroups. Our experiments suggest that there is very little potential to increase the social ties between different groups by changing the environment that leads to contact between students. Segmentation by race or background appears to be mainly driven by preferences to form friendships conditional on meeting, and less by differences in the probability of meeting. We also simulate the effects of an increase in the number of minority students. The result is more segmentation of the minority group in question. However, the fraction of minority friends of other races increases.

Our paper contributes to the existing literature along several dimensions. Much of the existing literature that studies the effects of one's academic peers exploits random assignment, such as the assignment of roommates (e.g. Sacerdote, 2001). However, most of a student's peers are not assigned randomly but are chosen through a process of network formation. We adapt methods from the networks literature to model the process by which a student forms her peers. The social networks literature has developed models of network formation with the goal of specifying parsimonious models that explain certain features of observed social networks such as clustering and small-world effects. Our goal is somewhat different. We seek to understand how the process of network formation contributes to the segmentation of heterogeneous individuals.

Existing social network models do not incorporate heterogeneity in agents, and therefore do not generate the observed segmentation. Therefore, we add heterogeneity in preferences and environment to evaluate education policies aimed at reducing social network segmentation.

In section II we describe our data, document the structure of the networks, and analyze associations between individual characteristics and friendship formation. In section III, we present a model of network formation and use it to simulate counterfactual networks in section IV. In Section V we show that a student's academic and social outcomes are related to her friends. Section VI concludes.

II. Description of Social Networks

A. Data from Facebook.com

Our data on social networks are from the Facebook.com website. The online student directory Facebook.com was conceived by undergraduate students at Harvard in February 2004. In spring 2004, the Facebook expanded beyond Harvard to other Ivy League schools and by fall 2004 Facebook.com had added websites for several hundred colleges and universities around the country. In July 2006, Facebook.com was the seventh most visited website in the United States. To participate on the Facebook, students must sign up using an official university email address, ensuring that they are members of the campus community. The Facebook allows students to set up a profile page which includes a picture, name, gender, high school, major, classes taken, political orientation, music tastes, hobbies and other interests as well as any musings the student wishes to share.[9] Students registered on the Facebook can browse the profiles of other students at their university.

Students use the website to share information and stay in contact with each other. Also, students can use the Facebook to find contact information for other students in one of their

[9] Facebook has changed features over time, so some current features were not available at the time that our data were collected.

classes. Facebook.com administrators claimed that many students log on almost every day and 9 out of 10 do so every week during the period when our data were collected.

The profiles of the students also contain a list of 'friends'. A Facebook friendship is formed if student A sends a friendship request via the website to student B and student B accepts A's friendship invitation. Student A appears as a friend on B's Facebook profile and vice versa. We use these friend connections as a proxy for a student's social network. These friends are likely to include not only close friends but also the "weak ties" that Granovetter[10] describes as being important for information transmission.

What Do 'Facebook-Friendships' Measure?

We conducted informal surveys about the nature of Facebook friendships in several undergraduate classes at Texas A&M. The students describe their Facebook friends as acquaintances made at school or social activities. Students say they would be willing to help most of their Facebook friends with a homework assignment. We also can provide slightly more formal evidence that Facebook friendships measure interaction on campus. After we collected our data, Facebook added an additional feature that allows students to self-report how they met each of their friends. Using a sample of this information for Texas A&M, we found that the main channels of meeting friends were being co-members of a school organization (26%), meeting through another friend (16%), attending the same high school (14%), and taking a course together (12%). Very few friendships appear to be merely online interactions (0.4%). We provide further evidence in section V that a student's friends, as measured by Facebook, are associated with educational outcomes.

---

[10] According to Granovetter's (1974) "Strength of Weak Ties" thesis, people often obtain better access to job information from those with *weak* ties, because those individuals operate in different social circles and are likely to have access to different sets of information. Granovetter cites evidence that professional, technical and managerial workers learned about new jobs through weak ties (27.8%), strong ties (16.7%), and moderate ties (55.6%).

Summary Statistics

We have a snapshot of data from Facebook websites at 10 Texas universities from January 17, 2005. The administrators of Facebook.com provided us with data on all student profiles at Rice, the University of Texas, Texas A&M University, Baylor, Texas Tech, Texas Christian University, Southern Methodist University, the University of North Texas, UT-Arlington and Texas State University.[11] Table 1 describes the undergraduate student body of the 10 universities. The schools are ordered by the date at which the Facebook was established on campus. Rice has the oldest and Texas State the youngest Facebook network. The two major public schools are the University of Texas and Texas A&M University while Rice, Baylor, Southern Methodist and Texas Christian are private. With the exception of Rice, Baylor and Southern Methodist, the student bodies of the schools in our sample consist almost exclusively of students from Texas. Whites comprise a strong majority of the student population including 60% at the University of Texas and 82% at Texas A&M. The largest minority group tends to be Hispanics (10% at Texas A&M and 15% at University of Texas) followed by Asians (3% at Texas A&M and 17% at University of Texas) and blacks (2% at Texas A&M and 4% at University of Texas).

As shown in the last row of Table 1, a large fraction of students are registered on the Facebook website at the time our sample was drawn. 80% of undergraduates at Rice, 40% at University of Texas and 44% at Texas A&M are included in our Facebook sample[12]. Altogether, our Facebook sample contains 65,104 undergraduate students.

In Section II.B we analyze the Facebook networks at all 10 Universities. In section II.C we match the Facebook data to additional student-level data from administrative records at Texas

[11] We thank Dustin Moscovitz for his assistance in providing us with the data.
[12] The last three schools in Table 1 (University of North Texas, UT-Arlington and Texas State University) were recent additions to Facebook when our data were collected, and fewer students participated in the Facebook in January 2005.

A&M. This allows us to look at the predictors of friendships in more detail. We also use the additional information to address selection issues.

B. Social Networks at 10 Texas Universities

First, we document the characteristics of the networks at all 10 universities. We show that the Facebook network exhibits characteristics common to social networks, and that the networks are strongly segmented by race, cohort, major, and political orientation. The Facebook does not ask students to report their race. To obtain the races of the students, we had 6 undergraduate research assistants classify the pictures on the Facebook profiles by race. The race categories used in this classification are: White/Hispanic, Black and East Asian.[13] The upper half of Table 2 shows the composition of the resulting networks. This sample is slightly more female and contains somewhat fewer minority students than the total student population at the 10 Universities.

Network Structure

A vast literature in sociology, mathematics and computer science provides an array of different tools to characterize networks. In order to present some of these measures, we need to introduce some notation.[14] We consider a campus with n students, or in the terminology of network analysis, a network with n nodes. Students i and j can be friends with each other, in which case the nodes i and j are linked or connected. This relationship is symmetric; if student i is a friend of student j, then student j is also a friend of student i. The friendships between students are recorded in the symmetric $n*n$ matrix $\boldsymbol{g}$. If student i and student j are friends, the corresponding elements of the friendship-matrix $\boldsymbol{g}$ are equal to one, $g(i,j)=1$ and $g(j,i)=1$. Otherwise, the elements of $\boldsymbol{g}$ are equal to zero.

[13] Each picture was evaluated by two research assistants. We only include students in our analysis if both research assistants' race evaluation coincided. As a check, we compare the race evaluation by the RAs and official race information for Texas A&M; in the vast majority of cases the classifications coincide (see Table A1 in the Appendix).
[14] The presentation here is based on Jackson (2006). For other ways to characterize networks, see Newman (2003) and Wasserman and Faust (1994).

One measure of the cliquishness of a network is the *cluster coefficient*. It captures the fraction of the friends of a given individual who are friends with each other. The literature considers different ways of calculating this measure. We follow Jackson and Rogers (forthcoming) and define the *total cluster coefficient* as:

$$C = \frac{\sum_{i:j\neq i,k\neq j,i} g_{ij}g_{jk}g_{ik}}{\sum_{i:j\neq i,k\neq j,i} g_{ij}g_{jk}}$$

The *cluster coefficient of an individual* is defined by:

$$C(i) = \frac{\sum_{j\neq i,k\neq j,i} g_{ij}g_{jk}g_{ik}}{\sum_{j\neq i,k\neq j,i} g_{ij}g_{ik}}$$

The average individual cluster coefficient is not necessarily equal to the total cluster-coefficient.

According to Jackson (2006) and Newman (2003), social networks are characterized by a number of common characteristics. The degree distribution (the distribution of the number of friends) is right skewed and has fat tails. Social networks tend to be cliquish and exhibit a cluster coefficient that cannot be explained by random formation of links. In a randomly generated network with many nodes and few connections, the cluster coefficient equals the probability that two nodes are connected and is close to zero. Newman (2003) and Jackson (2006) report cluster coefficients ranging from .09 to .45 for co-authorship networks in different academic disciplines; Goyal et al. (2006) report cluster coefficients from .16 to .20 among co-authors in economics. Newman also reports a cluster coefficient of .2 for a network of actors, where a link is established when 2 actors co-star in the same movie. Newman (2003) reports that social networks exhibit positive degree correlations -- nodes with many (few) links are connected to other nodes with many (few) links. Social networks also tend to exhibit a negative correlation between the individual cluster coefficient and the number of links of a node.

The lower half of Table 2 shows characteristics of the Facebook networks at the 10 universities.[15] It can be seen that the standard features of social networks are exhibited. The average number of friends ranges from 17.2 at the UT-Arlington to 62.9 at SMU. This can be partially explained by the date that Facebook started on each campus. The variance of the number of friends is closely associated with the mean; it ranges from 17.7 at UT-Arlington to 50.8 at Baylor. The number of friends is clearly right-skewed at all 10 universities. The skewness ranges from 1.06 at Rice to 2.28 at U North Texas.

All 10 networks are clustered. The cluster coefficient ranges from 0.17 at Texas A&M to 0.27 at UT-Arlington. Larger networks tend to have a smaller cluster coefficient[16]. The average individual cluster coefficient is in general slightly higher than the total cluster coefficient. The degree correlation is always positive -- it ranges from .22 at Rice to .58 at Baylor. The degree-cluster correlation is negative for all schools except the last three which were the youngest Facebook networks.

Segmentation of the Social Networks

Tables 3 and 4 show that the friendship networks at the 10 Texas universities are segmented by race, major, cohort, and political orientation. A variety of definitions and measures of segmentation, or segregation, have been proposed in the literature (see Echenique and Fryer (2005) and Newman (2003)). We compare the probability that two members of a subgroup are friends, to the probability that two random students are friends. This measure of relative segmentation is independent of the size of the two different groups. The upper part of Table 3 presents the fraction of pairs of students who form friendships conditional on the racial compositions of the pairs. In the middle part of the table, these probabilities are normalized

---

[15] We also have analyzed the networks by gender. The female-only networks tend to be bigger than the male-only networks. The higher number of average friends for females is associated with a higher variance of the number of friends. Otherwise, the networks for males and females look relatively similar.

[16] The cluster coefficient is directly related to the network density (the probability that any two nodes are connected). All else equal, denser networks have higher cluster coefficients. In a completely random network the cluster coefficient is given by the probability that any two students are friends. In a large network this is basically zero. In a small network this number is larger.

relative to the probability that two students of any race form a friendship. We call this number is the relative probability of friendship. For example, the relative probability of friendship for blacks is given by:

$$\text{Relative Probability of Friendship (black\&black)} = \frac{\dfrac{\text{Number of pairs of blacks who are friends}}{\text{Total number of pairs of blacks}}}{\dfrac{\text{Number of pairs of any students who are friends}}{\text{Total number of any pairs}}}.$$

It can be seen that students of the same race are more likely to form a friendship than students of different races. Most students are white/Hispanic and the probability that two white/Hispanic students form a friendship is similar to friendship formation of any two random students. Two Asian students are 1.59 (at UT Arlington) to 7.42 (at A&M) times more likely to be friends than any two random students. For pairs of blacks, this ratio ranges from 5 (at U of North Texas) to 16.5 (at A&M).[17] The relative probability of friendship is smaller than one for cross-race pairs.

The actual social environment of an individual is determined by the likelihood of forming a friendship with certain types *and* by the composition of the student body. The fraction of black friends of a black student depends on their relative probability of friendship formation and the share of blacks in the entire student population:

$$\begin{aligned}&\text{Fraction black friends of black student}\\ &= \text{Relative Probability of Friendship (black\&black)} \,*\, \text{(share of blacks in population)}.\end{aligned}$$

In the lower part of Table 3 we illustrate the resulting absolute segmentation. If friendships were formed randomly, the distribution of characteristics among the friends of any

[17] The segmentation by race is more pronounced for smaller minorities and at bigger institutions. Possible explanations are that smaller minorities stick together; and that a larger absolute number of students facilitates segmentation, as the number of minority students with specific interests increases. Future work could explore these conjectures.

subset of students should equal the distribution in the population, i.e. the fraction of Asian friends of Asian students should equal the fraction of Asians in the entire population. At all universities and for all races, students have a higher fraction of friends from their own race than implied by random assignment. For example, 13% of the students in our sample from Rice are Asian, but 30% of the friends of Asian students are Asian. 25% of the friends of blacks at Rice are Black while blacks comprise only 5% of the student population. While students have disproportionately many friends of the same race, it is also true that students mix across races. Students at more diverse universities have more diverse social networks. For example, white students at institutions with a large share of minorities tend to have more minority friends than whites at more homogeneous institutions.

Table 4 documents segmentation by major, cohort, and political orientation. Students have at least twice as many friends from the same major than random friend assignment would generate. Two students in the same cohort are about twice as likely to be friends, as two random students. The further apart the cohorts of two students are, the less likely these two students are friends, i.e. a freshman and a sophomore are more likely to form a friendship than a freshman and a junior. At all schools, self-reported conservatives have disproportionately many conservative friends and liberals disproportionately many liberal friends, however this segmentation is weaker than the race segmentation for minorities.

C. Friendship formation at Texas A&M

The 10 Facebook networks described in section II.B. are all segmented by race, major, cohort, and political orientation, and they all exhibit standard features of social networks.

From now on we focus on one of these networks, Texas A&M. For this university we have additional information about the students' characteristics. We link data from the Facebook, to administrative data from the Texas A&M registrar's office.[18] The additional administrative

[18] The students are linked by name and date of birth with a match success rate of 90%.

data include the academic record of the students (i.e. major, grade point average), race, the dorm a student lives in, membership in sororities and fraternities, information about parental background, SAT scores and high school information. We use administrative data on race rather than the visual race categorization used above. This allows us to distinguish Hispanics, who are the largest minority at Texas A&M.

In order to evaluate sample selection, we also obtained summary statistics of these variables for students not in the Facebook. Table 5 shows summary statistics for the students in the Facebook who were successfully matched to administrative records and summary statistics for the overall student population at Texas A&M. The two samples are very similar along most dimensions including GPA, SAT, high school percentile, and athletic participation. The Facebook tends to be slightly more popular among female students and among younger students. The latter feature explains the higher fraction of students living in a dorm in the Facebook sample. Members of fraternities / sororities (greek) are overrepresented in the Facebook. Two minority groups are slightly underrepresented -- blacks (2.3% of the Facebook population vs. 2.9% of the overall population) and Hispanics (11.4% of the Facebook population vs. 12.0% of the overall population). Students in the Facebook are slightly more likely to have college-educated parents and to come from a high income household.

We construct a sample that contains the 7,719 students in the Texas A&M Facebook network for whom we have complete data on race, demographics, family background, SAT scores, GPA and college activities.

We consider all pairs of students (i.e. $\frac{N(N-1)}{2}$ sets of possible friendship pairs) and quantify the relationship between their characteristics and the formation of friendships. We estimate a linear probability model of the form:

$$Friends_{ij} = f(X_i, X_j, \varepsilon_{ij}; \beta) \text{ for all } i \neq j$$

where $Friends_{ij}$ is an indicator of whether two students are Facebook friends and X_i, X_j are characteristics of the two students. We do not view this evidence as causal but merely as an analysis of the factors that are good predictors of friendship. The results are shown in Table 6.[19] When we condition on none of the students' characteristics, the probability that any two students are friends is 0.34 percent. Such a small baseline rate is not surprising for a large university -- the chance the any two random students have social contacts is small.

In the first column, we analyze the extent to which the race of students *i* and *j* serve as predictors of friendship. As seen above, students of the same race are more likely to form friendships. Both students being African American and both being Asian significantly increases the probability of being friends. Two students who are black are 17 times more likely to be friends than two students chosen at random (i.e. (0.0551+0.0036)/0.0034). Two Asian students are 5 times more likely to be friends. If both students are Hispanic or White, the probability of being friends increase to .53 and .36 percent, respectively. The probabilities that pairs consisting of a white and a minority student form a friendship are around .25 percent, lower than the baseline .34 percent. The differences in the same race coefficients are statistically different from one another.

In the second column, we test for associations between friendship and the students' gender, year in school, and whether the students went to the same high school. Not surprisingly, having attended the same high school increases the likelihood of being friends at college. Students in the same year of school are more likely to be friends and larger differences in years reduce the likelihood of friendships. Being of the same gender is a weak, but positive predictor of friendship.

[19] Table 6 considers all possible pairs for both male and female students. In unreported regressions, we estimate the equivalent results for male-only, female-only and cross gender pairs. We find that the factors that predict friendships are in the same order of magnitude across all gender combinations.

The third column indicates that family background is associated with friendships in college. Students from families with similar income levels are more likely to be friends. In addition, students are more likely to be friends if each has at least one parent with a college education. Friendships are less likely between students with no college educated parents and between students in which one has college educated parents and the other does not. This suggests that social networks are at least partially segmented along socio-economic lines. Finally, students are more likely to be friends if they share the same political orientation.

Institutional factors that influence the likelihood that students will meet each other are strongly associated with friendship. As seen in the fourth column, being in the same dorm leads to about a 13 fold increase in the probability of being friends relative to two randomly chosen students. Other institutional factors, such as being in the same major or the same college, also increase the likelihood of friendship, however these effects are an order of magnitude smaller than living in the same dorm.

Friends also appear to be somewhat segmented by ability, as seen in column 5. Two students are slightly less likely to be friends relative to the baseline if they have SAT scores or college GPAs that differ substantially. However, the magnitude of these factors is smaller than those for the race and institutional factors.

Finally, we analyze the association between campus activities and the probability of being friends. Our measures of activities are participation in intercollegiate athletics (2.5% of students in sample), a greek organization (11.6% of students in sample), and the Corps of Cadets (1.8% of students in sample). Not surprisingly, students are more likely to be friends if both participate in the activity relative to the case of neither student participating. Similarly, if only one student participates in the activity, students are less likely to be friends than if neither participates. This segmentation is strongest for athletes and Corps members.

In the seventh column, we include all sets of characteristics as predictors. Many of the coefficients in this model are very similar to their counterpart in the model with fewer covariates.

In particular, the coefficients for race are robust to controlling for demographics, ability, dorm, major and activities. The fact that adding covariates does not significantly change the race coefficients suggests that the observed social network segmentation by race does not merely reflect different institutional channels of meeting such as major, athletics or dorm. Rather it suggests that there are tastes for characteristics that are correlated with race that affect the probability of becoming friends. However, the importance of similar socioeconomic backgrounds diminishes when controlling for other common characteristics.

These results suggest that the factors most strongly associated with friendship are sharing the same high school, same dorm, same race for minority students, same campus organizations, same major/college, same political orientation, being from the same cohort, and to a lesser extent sharing similar parental background characteristics.[20] We incorporate these insights into our model below.

Note that in each of the models discussed above, the R^2 is low. This is not surprising. There are many unobserved characteristics, tastes, and coincidences that determine friendship formation. We can illustrate the importance of one of these additional factors -- having common friends. The last column of Table 6 shows the results of the linear probability model when we include the number of common friends as an additional regressor. Common friends are a good predictor for the existence of a friendship -- the R^2 increases from below .04 to almost .25. Moreover, conditioning on the number of common friends changes the importance of other characteristics, such as common race, for the formation of friendships. The coefficient for both black drops by two-thirds and the coefficients for both Asian and both Hispanic drop by one-third and one-quarter, respectively. It is difficult to interpret the meaning of these changes, as the formation of friendships and the determination of friends of friends are the outcomes of the same process.

[20] These results are consistent with the findings of Marmaros and Sacerdote (2006). They report that physical distance of residence and cohort are two important institutional factors that determine the interaction between students. Ward (2004) also studies the effect of distance on interaction.

This suggests that if we want to assess effects of changes in any of the variables determining friendship formation, it is not sufficient to use the results of the linear probability model to predict changes in the composition of friendships. Endogenous effects through friends of friends may magnify effects of a changed environment. The probability that student *i* and *j* meet is a function of whether they are both friends with student *k*, so characteristics of student *k* also affect the probability that *i* and *j* are friends. Therefore, we proceed by building a model of friendship formation.

III. Model of Social Network Formation

We seek to understand the process of network formation and quantify the importance of different determinants of the network, while taking endogenous network effects into account. A model of social network formation makes it possible to evaluate policies that alter social interactions in college. For example, such a model could allow university administrators and higher education officials to evaluate the effect of increasing the number of ethnic minorities, admitting students from different parts of the ability distribution, or changing freshman dorm assignments.

Our model combines a stochastic meeting process and choices by individuals based on their preferences. Like in Jackson and Rogers (forthcoming), the meeting process consists of random encounters and introduction to friends of friends. We add heterogeneous agents and a simple preference structure. We maintain the assumption that agents do not take existing or future links into account when choosing to form a link.[21] We calibrate our model to fit data on actual networks and conduct counterfactual experiments.

---

[21] This model of network formation is rudimentary in several dimensions. The decision to form a friendship conditional upon meeting is based solely upon the characteristics of the two students. The network formation literature has more developed theoretical models in which a network is the equilibrium outcome of a noncooperative game. For a good survey of the literature on the theory of network formation, see Jackson (2005). In some of these models, the decision to form a link is based upon the position of each

A. Mechanics of the Model

The starting point of our model is a completely unconnected network. No friendships have been formed and all elements of the friendship-matrix $\boldsymbol{g}$ are equal to zero. We conceptualize a friendship between students i and j as the outcome of two events: (1) two students meet with some probability, and (2) conditional upon meeting, students choose whether or not to form a friendship. Students i and j meet each other with a probability $p_{ij}(Z_i,Z_j)$, which is a function of observable features of each student's institutional environment $Z_{i/j}$ (e.g. living in the same dorm or being part of the same cohort). In addition, students meet other students through their existing friends.

After two students meet, they decide whether they like each other. This decision depends on one another's characteristics, some of which are observable ($X_{i/j}$) and some of which are unobservable ($u_{i/j}$) to the analyst. Denote $U_{ij}(X_i,X_j,u_i,u_j;\beta)$ the utility student i derives from being friends with student j, and c_i the marginal cost of friendship to student i (e.g. the time cost of a friendship). Because friendship is mutual, we model friendship as:

$$g(i,j) = \boldsymbol{I}(U_{ij}(.) \geq c_i) \cdot \boldsymbol{I}(U_{ji}(.) \geq c_j) \quad \text{for any } i,j \text{ that meet}$$

$$\equiv \boldsymbol{I}(f(X_i,X_j,u_{ij};\beta) > 0)$$

where $\boldsymbol{I}(.)$ is the indicator function.

In the second line, we represent the joint choice to be friends with a reduced-form mutual friendship function f. The parameters of this function (β) represent tastes for the observed parameters as well the marginal cost of friendship. u_{ij} is a reduced-form representation of students i and j's unobservable tastes for one another.

node in the network and the existing links of the network. Other work has studied properties determining whether a network is stable and if any player would want to sever a link given the current network configuration.

The characteristics in X that affect the mutual friendship function are race, parental education, SAT score, and political orientation. The functional form used in the simulation is given by:

$$
\begin{aligned}
&f\left(X_i, X_j, u_{ij}; \beta\right) = \\
&\beta_0 + \beta_{WW} I\left(race_i = race_j = white\right) + \beta_{BB} I\left(race_i = race_j = black\right) \\
&+\beta_{HH} I\left(race_i = race_j = hispanic\right) + \beta_{AA} I\left(race_i = race_j = asian\right) \\
&+\beta_{par\_edu} I\left(parent\_edu_i = parent\_edu_j = both\_coll\right) \\
&+\beta_{cons} I\left(conservative_i = conservative_j = 1\right) \\
&+\beta_{skill} I\left(SAT_i > 1200 \,\&\, SAT_j > 1200\right) + u_{ij}
\end{aligned}
$$

where u_{ij} captures the joint effect of the unobservable characteristics of i and j. The β coefficients capture tastes for similar characteristics. We allow for similar characteristics to increase or decrease the likelihood of friendship, but we restrict the taste parameters to be equal across students. In the calibration of the model, u_{ij} is simulated with independent random draws from a normal distribution[22]. The mean and variance are normalized to zero and one. The magnitudes of the other parameters in the function are relative to the variation in the random component.

The meeting is modeled in different stages. First each student meets every other student in the university with probability p_{init}, and this probability is chosen to generate on average c_{init} meetings per person. Next each student meets each other student from the same college with a probability p^i_{COLL}, chosen to generate an average of c_{COLL} meetings per person.[23] Students of the same cohort meet each other with probability p_{Year}. The students living in a dorm meet each other

[22] To check whether the results are sensitive to the independence assumption we recalibrate the model while imposing a connection between u_{ij} and u_{ik}, and u_{jk}. (See Appendix)

[23] Texas A&M has 10 different academic colleges, e.g. Liberal Arts, Engineering, or Architecture. Because some colleges are larger than others and a student is less likely to meet any other student in the college if the college is large, we allow the probability p^i_{COLL} to vary by individual. The probability varies in such a way that every student meets c_{COLL} students on average from their college, independent of the size of their college.

student living in the same dorm with probability p_{DORM} . The rule $\boldsymbol{I}(f(X_i,X_j,u_{ij};\beta) > 0)$ is used to decide whether a meeting through any one of these channels results in a friendship. The order of these channels does not matter.

After meeting a set of initial friends, students meet the friends of their friends. This process is motivated by the clusteredness of the networks documented in section II. A model with different probabilities of friendship formation can generate the segmentation observed, but cannot produce clusteredness within subgroups[24]. The process of meeting friends of friends can magnify any effects of the institutional environment on friendship formation. Each student meets each friend of her friends with probability p_{frofr}[25]. The friend of friend meeting process is repeated S times[26]. Again $\boldsymbol{I}(f(X_i,X_j,u_{ij};\beta) > 0)$ is used to decide whether a meeting results in a friendship.

These multiple rounds of meeting and consequent decision whether a friendship is formed result in a friendship matrix $\boldsymbol{g}$. We calculate features describing this simulated network. In section IV, we calibrate the 14 parameters so as to fit 14 moments of the simulated network to 14 moments of the actual network at Texas A&M. The moments are: the mean, variance and skewness of the number of friends; the cluster coefficient; the fraction of friends from the same college, same dorm or same cohort; the fraction of friends who are the same race for whites, Hispanics, Asians and blacks; the fraction of friends who are high SAT scorers for high SAT

[24] If both i and j are friends with k, the friends of friend mechanism increases the probability that i meets j. A closed triangle between i, j, and k contributes to clusteredness. The meeting of friends of friends also generates the positive degree correlation. Consider two nodes one (i) has few friends after the initial meeting stage another one (j) has many friends after initial meeting. Both i and j introduce their friends to each other, since j has more friends than i, the friends of j will meet more people than the friends of i. The result is that j will be friends with people with a lot of friends and i will be friends with other people with few friends; we observe a positive degree correlation.

[25] We recalibrate the model while relaxing this assumption (see the Appendix). The counterfactual simulations based on the alternative specification generate very similar results as in the original specification.

[26] It is possible that the same two people meet multiple times. If they like each other initially, they form a friendship and stay friends. If they do not like each other the first time, they will not form a friendship at subsequent meetings, as well. We assume that whether two people like each other is determined once and does not change over the course of the network formation.

scorers; the fraction of friends of the same parental education level; and the fraction of conservative friends of conservatives.

The mechanics of the model implies that all parameters of the model affect all moments. But it is possible to illustrate the relationship between the different parameters and a given moment to illustrate how the moments are determined. The number of total friends is directly related to the number of students randomly met. The channel of meeting friends of friends generates the variance and skewness of the distribution of the number of friends, as well as, the clusteredness. Hence these three moments are directly related to the number of cycles of meeting friends of friends (S), the number of friends met in each cycle (c_{frofr}), and the probability of forming a friendship conditional on meeting, captured by the intercept (β_0) in the function $f\left(X_i, X_j, u_{ij}\right)$. The fraction of friends in a similar environment is directly related to the probabilities of meeting people in that environment (c_{COLL}, p_{YEAR}, p_{DORM})[27].

The fraction of friends with the same characteristics implies values for the importance of sharing these characteristics when deciding to form a friendship, i.e. the parameters β_{WW} - both white , β_{BB} - both black, β_{AA} - both Asian, β_{HH} - both Hispanic, β_{HiSAT} - both high SAT score, β_{par_edu} - same parental education, and β_{cons} - both conservative.

B. Assumptions and Exclusion Restrictions

The model postulates that the probability that two students meet is determined by specific institutional factors (academic college, dorm, cohort). Preferences for friendship conditional upon meeting are determined by specific observable characteristics (race, parental education, political orientation, and academic ability). We assume that unobserved determinants of tastes are uncorrelated with institutional meeting channels. If this assumption is violated, the model

[27] Recall that the average number of students from the same college that an individual meets is given by c_{COLL}, while the probability of meeting each individual from the same cohort / dormitory is given by p_{YEAR} and p_{DORM}.

will yield biased parameters of the effects of institutional variables. For example, suppose two political science majors share an (unobserved) interest in campus politics. The shared major is an institutional meeting channel that is correlated with the unobserved shared taste for politics, and therefore affects the probability of becoming friends conditional upon meeting. If two political science majors are more likely to be friends conditional upon meeting, we bias upwards the parameters that capture the causal effect of sharing the same major.

Equivalently, we assume that unobserved determinants of meeting are uncorrelated with observable taste characteristics. For example, we rule out that two high achieving students in the same dorm are more likely to meet through institutional meeting channels than a high and a low achieving student in the same dorm. In particular, this assumes that the university does not have unobserved meeting channels that affect the probability of meeting but are correlated with our measures of taste (e.g honors classes, student associations for certain ethnicities).

These assumptions are motivated by the reduced-form regressions in section II. The coefficient estimates of the institutional variables (e.g. same college, cohort) are fairly robust to the inclusion of a variety of covariates on ethnicity, family background, and ability. If the additional covariates pick up any of the unobserved heterogeneity, the robustness of these regressions suggests that the bias may not be severe. In addition, the coefficient estimates of same race are robust to the inclusion of institutional variables. This suggests that the observed institutional variables are related to friendship in a manner that is largely independent of race, and supports the validity of our key identifying assumption.

IV. Results and Simulations

Now, we parameterize the model presented in Section III, assess the model fit, and perform several simulations to illustrate the mechanics of network formation and evaluate the consequences of policies.

A. Model Calibration

We calibrate[28] the 14 parameters of the model to fit 14 moments of the simulated network to 14 moments of the actual network at Texas A&M. We calculate the data moments using the Facebook network of 1930 students randomly drawn from the sample introduced in section II.C..[29] The 14 data moments are displayed in column one of Table 8. We use the characteristics of the 1930 individuals and simulate a network by applying the network formation mechanism presented in section III.

We adjust the parameters in subsequent simulations to minimize the difference between the features of the simulated network and the features of the network at A&M. The simulations are based on random draws for who meets whom and the elements of u_{ij}. For each set of parameters the network features of the simulated model are calculated by averaging over 100 simulated networks with different random draws. The 100 sets of random components are kept constant for each different set of parameters. The parameters found by this process are displayed in column 2 of Table 7.

The simulated network is generated by meeting on average 6.1 random students, 2.1% of all students in the same cohort, and 4.6 students from the same college. Students living in the same dorm meet each other with 35% probability. Conditional on meeting, whites have a very small preference for friendships with other whites (β_{WW}=.07). The preferences for same race friendships are much stronger for Hispanics (β_{HH}=.40) , Asians (β_{AA}=.85), and especially Blacks

---

[28] We do not estimate but calibrate the model, which means that the resulting parameters cannot be used for testing or to construct confidence intervals.

[29] Due to computational limits, we have to restrict ourselves to a subset of the 7719 students in the full sample. Therefore, the reader should interpret the counts we present below as corresponding to a "scaled down" network. We drew different samples of 1930 students. The characteristics of the network remain essentially unchanged. While students with many friends will loose more friends due to sampling than students with few friends, the relative distribution of the number of friends is not affected by using a random sub-sample.

(β_{BB}=2.10). The preferences for friends with similar SAT scores, parental background or political orientation are less pronounced than the preferences for same race friendships among minorities.

Column 2 in Table 8 displays the features of the model simulated with the parameters shown in Table 7. Our model generates the features of the network. It matches the average number of friends and the variation in the number of friends. It also produces the right skewed distribution of the number of friends, as well as the clusteredness of the friendships.[30] The model matches the likelihood of forming a friendship conditional on sharing a similar environment or similar characteristics.

B. Assessing the Fit of the Calibrated Model

We can calculate a variety of other moments that did not enter the calibration using the parameters from Table 7. This allows us to assess the suitability of the model. In Table 8 we present additional moments for both the real world network and the simulated model. We calculate various metrics that condition on common cohort, college and dorm. We focus on whites and Hispanics because these are large groups in our sample; there are fewer Asians and blacks for whom our model is less precise. In the data, the fraction of same race friends of Whites and Hispanics is virtually unchanged when conditioning on common cohort or college. Our model generates this pattern. When conditioning on living in the same dorm, the model closely matches the fraction of same race friends for whites but overstates the fraction for Hispanics.

We find that our model predicts other key features of social networks. The model predicts an average individual cluster coefficient (0.16) that is relatively similar to the actual data (0.12). Also, the model correctly predicts the high positive degree correlation observed in the sample (0.61 vs. 0.49). Finally, the model predicts a positive degree-cluster correlation which is

[30] The features are not perfectly fitted. One reason is that the number of cycles of meeting friends of friends is an integer rather than a continuous parameter.

consistent with the data, but the model underpredicts its magnitude (0.15 vs. 0.24). Although these other moments are by no means perfectly fitted, these results suggest that our model captures key underlying mechanics of network formation.

C. Counterfactual Experiments

To simulate counterfactual networks, we use the parameters obtained above but change various elements of the network formation process. This allows us to evaluate possible policy changes. We assume that the parameters of the model are not affected by the policy changes considered. This can be justified by the fact that any actual changes are most likely only marginal. However, for more substantial policy changes our approach is subject to the Lucas critique.

The first counterfactual is purely random friendship formation. This provides a benchmark against which to compare other processes of network formation. The parameters used for this simulation can be found in column (3) of Table 7. Each student meets each other student with the same probability independent of their environment, and the probability of forming a friendship conditional on meeting does not depend on any characteristics of the students. The average number of random friends is chosen to make this random network comparable to the actual network and the simulation of the full model. The features of the resulting network are presented in column (3) of Table 8. The network does not exhibit the common features of social networks. The number of friends varies less, is not very skewed, and there is almost no clustering. Under purely random friendships, the fraction of friends with certain characteristics reflects the share of the total population, and there is no segmentation.

The second counterfactual experiment shows that meeting friends through other friends can magnify certain measures of segmentation and mitigate others. In this simulation, we "turn off" the friends of friends meeting channel -- students meet with probabilities that vary in school environment and they have preferences for friend characteristics, but they do not meet friends of their friends. The parameters for this counterfactual simulation are presented in column (4) of Table 7. In column (4) of Table 8, we present the features of the resulting network. The segmentation based upon dorm and cohort is larger, suggesting that the friends of friends channel facilitates becoming friends with students in other school environments. However, the friends of friends channel magnifies racial segmentation for blacks – 22% of the friends of blacks are black without the friends of friends channel while 33% are black when we allow for meeting friends through friends. This result illustrates the importance of modeling network-based meeting.

Column (5) of Tables 7 and 8 shows the parameters and resulting network features for the counterfactual of "random meeting". This counterfactual is intended to model the extreme case of a university eliminating any meeting channels that generate segmentation. Obviously, it would be impossible to eliminate all such channels, but a university could, for example, create a common set of core classes that students from all academic colleges must take. In this simulation, each student meets every other student with equal probability (i.e. the institutional meeting channels do not affect meeting probabilities), but students have the preferences that we estimate above and meet the friends of their friends.[31] We find that the variation and skewness of the number of friends decrease slightly but the cluster coefficient remains virtually unchanged. As expected, the disproportionate number of friends with a similar campus environment disappears. However, the segmentation by race, ability, political orientation and parental education largely persists. This is a potentially sobering result for university administrators. It suggests that university policies geared towards increasing the encounters between different groups of students have very limited ability to reduce segmentation in their students' social networks.

Next, we perform the "reverse" counterfactual. Rather than eliminating meeting channels and maintaining preferences, we consider the case of undiscriminatory preferences with existing meeting channels. The likelihood of forming a friendship conditional on meeting does not

[31] Mechanically, the average number of random encounters of each student is picked to generate the same average number of friends as in the original network.

depend on the characteristics of a person. The probabilities of meeting other students are the same ones as in the full model (i.e. the institutional channels affect meeting probabilities and students meet friends of friends). The probability of forming a friendship is chosen to match the average number of total friends. The parameters we use are in column (6) of Table 7 and results in column (6) of Table 8. We observe numbers of same race friendships that are very close to those that would arise from pure random friendship formation. This confirms the result that the segmentation according to race in the actual network is mainly driven by preferences and not by different meeting probabilities. This supports that the reason that, say, Hispanics have disproportionately many Hispanic friends is preferences. It argues against the alternative explanation that Hispanics meet disproportionately many Hispanics through channels such as major or dorm and these differences are then magnified through introduction to friends of friends.

We also can simulate the effect of an affirmative action policy that admits more students of a certain demographic profile. We choose to model the admission of more Hispanics who are a large and growing population in Texas. In column (7) we enact the policy experiment of doubling the population of Hispanic students. We do so by including each Hispanic student with all her characteristics twice in the simulation. We assume that preferences for race do not change. We find that Hispanics would have a much more racially segmented social network -- the share of Hispanic friends of Hispanics nearly doubles. However, the share of friends of a different race increases for whites, Asians and blacks. In particular, the share of non-white friends of whites increases from 15% to 24%. This implies that increasing the number of Hispanics would lead to modest increases in the racial diversity of interaction for non-Hispanics.

In column (8), we show the results for a counterfactual experiment where we introduce whites and minority students to each other. This policy experiment would correspond to intentional efforts by the university to facilitate interaction between students of different backgrounds (e.g. targeted introductions during orientation week). To perform this simulation, we include an extra meeting round, where each white student has a 1% chance of meeting each minority student and each minority student has a 1% chance of meeting each white student. This translates to each white student meeting 3.5 non-white students and each minority student meeting 15 white students.[32] The probability of forming a friendship conditional on meeting is still given by the preferences used to simulate the full model. Given our previous finding that preferences significantly affect friendship formation, we would expect this policy to have only limited effects. Indeed, we find that the diversity of social interaction only modestly increases. The result is an increase in the share of minority friends of white students from 15% to 23%. The total number of friends of minority students increases, but their share of same race friends decreases only slightly.

The results of the counterfactual experiments' effect on racial segmentation are summarized in Figure 1. The results suggest that changes in institutional policies have limited potential to increase inter-race interaction. Equalizing the probability of meeting (e.g. random dorm assignment and/or a common set of core classes) only negligibly changes the fraction of inter-race friendships for Whites, Hispanics, and Asians. A policy of targeted introductions of minorities to non-minorities has the largest impact on whites by increasing the number of non-minority friends, but has little impact on Hispanics and Asians. However, alternative institutional policies do have modest impacts on blacks. Targeted introductions increase inter-race friendships from 67% to 69% while equalizing the meeting probability increases the figure to 72%. This result is likely driven by the finding that the magnification effect of preferences through friends of friends is strongest for blacks. Thus, policies that change meeting probabilities can impact blacks' social network despite strong same-race preferences.

[32] It turns out that about 15% of all contacts of white students and 50% of all contacts of minority students are through this channel.

V. Evidence that Facebook Social Networks Are Associated With Educational Outcomes

As pointed out in the introduction, social contacts may affect many outcomes that are of interest to economists. When using data from an educational setting, many of these outcomes are not observable or have not manifested themselves yet. We do not observe how diverse interaction promotes the diversity of thought, how students' social networks affect their job search, or how their eventual productivity is affected by their connections. However, we are able to document some associations of academic and social outcomes (Foster (2006) points out that in student populations peer effects tend to be more robust in social outcomes than in academic outcomes.)[33] These outcomes might be interesting for their own sake, or can be viewed either as proxies for others outcomes or precursors for subsequent outcomes. Because we measure both endogenously and exogenously determined friends, we do not interpret any of these associations as causal.[34]

We regress a student's outcome on own-pre-treatment characteristics, friend pre-treatment characteristics, and contemporaneous friend characteristics. We study five student outcomes – GPA, drinking behavior, and participation in three types of organizations. GPA is a student's college GPA when the Facebook data were collected. 'Drinker' as an indicator that the student self-reported on their Facebook profile's "Interests" section words typically associated

[33] A variety of studies have examined the relationship between a student's peers in college and her outcomes along a variety of dimensions including academic performance, choice of major, choice of extracurricular activities, and social behavior. The existing literature contains an emerging set of studies that cleverly use random assignment of peers in college, for example, freshman year roommates or peers in their first year at a military academy. For example, see Sacerdote (2001), Zimmerman (2003), Kremer and Levy (2003), Winston and Zimmerman (2003), Boisjoly et al. (2004, 2004), Foster (2006), Stinebrickner and Stinebrickner (2006), Sacerdote and Marmaros (2006), Lyle (2007) and Carrell, West and Malmstron (forthcoming). Results have been mixed, and it appears that peer effects may be stronger determinants of social than academic outcomes. Recent work suggests that it is not well understood which peers generate such effects and via which mechanisms that peer effects manifest, suggesting that further theoretical modeling may be insightful (Foster, 2006).

[34] See Manski (1993) for a discussion of these identification problems.

with drinking.[35] We use Facebook profile information on "Jobs and Clubs" to identify students who participate in three types of organizations – volunteer, religious, and political. "Volunteer", "Religious", and "Political" are indicator variables for whether the student reports on the Facebook that she participates in a campus or community volunteer, religious or political organizations, respectively.[36]

A student's own pre-treatment characteristics include measures of ability (SAT score and high school percentile), family characteristics (parental income and education), high school characteristics, race, and gender. A student's friends' pre-treatment characteristics are average friend SAT score, high school percentile, and parental education. We also include a measure of each student's own-race segregation – the Spectral Segregation Index (SSI) – recently developed by Echenique and Fryer (forthcoming).[37] Finally, we include several measures of the average contemporaneous characteristics of a student's friends.

Results are reported in Table 9. The last group of variables in each regression is the contemporaneous friend characteristics. If these characteristics are strongly related to a student's outcome, as we find below, these results provide support for the claim that a student's outcomes are strongly related to the outcomes of her friends.

The first 2 columns report the relationships between a student's GPA and her friends' average characteristics. In column (1), we include the student's own characteristics and friends' predetermined characteristics. Students with higher GPAs are those with higher measures of ability (SAT and percentile rank in their high school), a college educated father, and those who

[35] For example, students who list an Interest in "beer", "liquor", "drinking", and "partying" are classified as a 'drinker'. 10% of students in our sample are 'drinkers' according to this measure.
[36] To measure if a student participates in a these types of organizations, we collected all self-reported organizational membership in the Facebook data. Then, two research assistants who are familiar with student organizations independently classified the organizations as being volunteer, political, religious, or other. We classify the organization if both research assistants agreed on the classification. The working definitions of volunteer, political, and religious that were given to the research assistants are available upon request. According to this measure, the percentage of students participating in volunteer, religious and political organizations is 9%, 8%, and 3%, respectively.
[37] The SSI allows researchers to calculate an individual's own-race segregation. Computer code for calculating the SSI can be found at: http://www.economics.harvard.edu/faculty/fryer/projects.html. We thank Kimon Ioannides for his assistance with calculating the SSI.

went to wealthier high schools with higher scores on standardized tests. In addition, students have higher GPAs when their friends have college educated fathers. Friends' ability measures – SAT and high school percentile – are not individually significant, but both are positively related to GPA and jointly significant at the 5% level. Column (2) adds measures of contemporaneous friend characteristics. Average friend GPA is very strongly associated with own GPA – an increase in one letter grade in average friend GPA is associated with about half a letter grade increase in own GPA. Friends' predetermined skill measures are no longer positive after controlling for friend GPA. Having more friends who are 'drinkers' is associated with a lower GPA but the effect is only significant at the 11% level. It is worth restating that these coefficient estimates should not be interpreted causally. However, these results do suggest that the determination of a student's GPA is not at all independent of the determination of her peers, and this is reflected in our Facebook data.

In column (3) of Table 9, we report estimates of a linear probability model of being a 'drinker'. Males from higher income families are more likely to be 'drinkers'. Students with friends with lower SAT scores and lower high school percentiles are more likely to be drinkers; the coefficients are not individually significant but are jointly significant at the 1% level. Finally, a student with more friends who are 'drinkers' is substantially more likely to be a 'drinker'. A one standard deviation increase in the fraction of drinker friends (10%) increases the probability of being a 'drinker' by 4%.

The final sets of columns investigate the participation in volunteer, religious and political organizations with a linear probability model. For each outcome, we regress an indicator variable for whether the student is a member of the type of organization on own/friend characteristics and the fraction of her friends who are also in that type of organization. Friends being in the same type of organization reflect at least two mechanisms: friends may have similar preferences to participate in certain activities, or friends may meet via the organization. In the first column of each model, we calculate the fraction of friends by excluding friends who are in the same organization. For example, for a student in the volunteer organization Habitat for Humanity, we only count friends who are in volunteer organizations other than Habitat. This is intended to avoid counting friends who meet through the organization, and primarily identify friends with similar preferences to participate in certain types of organizations. In the second column of each model, we include friends who are in the same organization, so this measure picks up friends with similar preferences as well as friends who may meet in the organization. One would expect the coefficient of "fraction of friends" to be larger when we include friends in the same organization.

Column (4) shows that students with more friends in other volunteer groups are no more likely to volunteer, but column (5) shows that having more friends in any volunteer group makes the student much more likely to volunteer. A one standard deviation increase in the fraction of friends who volunteer (10%) increases the likelihood that the student volunteers by 5.6% (=0.10 * 0.564). This is suggestive that volunteer groups serve as a meeting channel.

The association between own and friend membership is even stronger for religious organizations. Even when we include only friends who are in different religious groups, a student with more friends in religious organizations is associated with a substantial increase in the likelihood of being in such a group oneself. This suggests that students who are friends are likely to have similar underlying preferences for religious fellowship. When we include friends in any religious organization in column (7), the association is even stronger – a one standard deviation increase in the fraction of such friends (13%) is associated with a 9.9% (=.13*.78) increase in probability of joining a religious organization.

For participation in political organizations, the association between own and friend membership is weak. Under both measures of friend membership, an increase in friend membership is associated with higher own membership but the relationship is not statistically significant.

We emphasize that these relationships should not be interpreted causally. A variety of factors – selection into friendships, the reflection problem and endogenous and exogenous peer effects – make causal inference problematic. Nevertheless, we obtain estimates of associations between student outcomes and friend characteristics that are consistent with the existence of some peer effects. In addition, these results provide further evidence that "Facebook friends" are students' peers with whom they interact and influence in class, student organizations, the dorm or other campus activities.

VI. Conclusions

The data from the Facebook networks offer insights into the social networks that impact learning, information transmission, and labor market outcomes at the beginning of adulthood. The data provide a large-scale view of the social networks at universities of various sizes. These social networks exhibit many of the characteristics suggested by the network structure literature – clustering, positive degree correlation, and variance and skewness of the degree distribution. In addition, we quantify segmentation along racial and socioeconomic lines and document the diversity of interaction on university campuses.

Our model provides a methodology to analyze segmentation in social networks and decompose the contribution of both school environment and preferences to observed segmentation. We illustrate the role of connections though friends of friends in the formation of social networks. The model relies on several key assumptions. We assume that certain observable student characteristics affect the probability that two students meet but do not affect the probability of forming a friendship conditional upon meeting. Likewise, we assume that other characteristics affect preferences to form a friendship but do not affect the meeting probability. Although we cannot directly test these assumptions, the patterns exhibited in the reduced-form regressions in section II.C. are consistent with these assumptions.

Our findings offer a mixed message for university administrators who seek to create diverse social interaction on campus. On one hand, social networks exhibit only modest segmentation across some important dimensions. In the actual network, the fraction of friends with similar ability, parental education, and political orientation does not differ substantially from the fraction that would be generated by random assignment of friends. This suggests that diverse interaction does occur.

However, social networks are highly segmented by race, and this is present at schools ranging from small private institutions (Rice, SMU) to large public universities (Texas A&M and University of Texas). Moreover, our counterfactual simulations suggest that racial segmentation is largely driven by preferences rather than institutional features that affect meeting. Changes in university policies that affect student meeting channels appear to have limited ability to reduce racial segmentation. Therefore, policies aimed at increasing social integration need to somehow impact preferences.

Our network formation model does not allow for strategic interaction between agents. Rather, it provides a purely mechanical meeting process, which is assumed to be independent of preferences for friendship formation. The challenge for future work is to develop models that allow for more complex mechanisms of network formation, while being still empirically tractable.

References

Arrow, Kenneth J. and Ron Borzekowski. "Limited Network Connections and the Distribution of Wages", Federal Reserve Board, Finance and Economics Discussion Series: No. 2004-41, 2004.

Babcock, Philip. "From Ties to Gains? Evidence on Connectedness, Skill Acquisition, and Diversity", mimeo, UC Santa Barbara, 2006.

Betts, Julian R. and Darlene Morell. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects", *Journal of Human Resources*, 34(2), Spring 1999.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy and Jacque Eccles, "Empathy or Antipathy? The Impact of Diversity", mimeo, Harvard, 2004.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy and Jacque Eccles, "Peer Effects in Binge Drinking Among College Students", mimeo, Harvard, 2003.

Bok, Derek C. and William G. Bowen "The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions" *Princeton University Press*, 1998

Calvo-Armengol, Antoni and Matthew O. Jackson. "The Effects of Social Networks on Employment and Inequality", *American Economic Review*, 94(3), June 2004.

Calvo-Armengol, Antoni and Matthew O. Jackson. "Like Father, Like Son: Network Externalities, Parent-Child Correlation in Behavior, and Social Mobility", mimeo, 2006.

Carrell, Scott E., Frederick V. Malmstrom, and James E. West. "Peer Effects in Academic Cheating, *Journal of Human Resources*, forthcoming.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy. "The Small World of Investing: Board Connections and Mutual Fund Returns", NBER Working Paper, 2007.

Echenique, Federico and Roland Fryer. "On the Measurement of Segregation", Harvard mimeo, September 2005.

Echenique, Federico, Roland Fryer, and Alex Kaufman. "Is School Segregation Good or Bad?", *American Economic Review*, 96(2), May 2006.

Fontaine, Francois. "Why Are Similar Workers Paid Differently? The Role of Social Networks", IZA Discussion Papers 1786, Institute for the Study of Labor (IZA), 2005.

Foster, Gigi. "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism", *Journal of Public Economics*, Vol. 90, 2006.

Foster, Gigi "Making friends: A nonexperimental analysis of social pair formation" *Human Relations* Vol. 58(11), 2005

Fryer, Roland and Paul Torelli. "An Empirical Analysis of 'Acting White'", mimeo, Harvard, 2006.

Grutter v. Bollinger *United States Supreme Court* (02-241) 539 U.S. 306 (2003).

Goyal, Sanjeev, Marco J. van der Leij, and Jose Luis Moraga-Gonzalez. "Economics: An Emerging Small World?", *Journal of Political Economy*, 114(2), 2006.

Granovetter, Mark S. "The Strength of Weak Ties", *American Journal of Sociology*, 78(6), 1973.

Granovetter, Mark S. *Getting a Job: A Study of Contacts and Careers*, University of Chicago Press, 1974.

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely. "What Makes You Click? – Mate Preferences and Matching Outcomes in Online Dating", mimeo, University of Chicago, 2006.

Ioannides, Yannis M. and Linda Datcher Loury. "Job information Networks, Neighborhood Effects and Inequality", *Journal of Economic Literature*, 42(4), December 2004.

Ioannides, Yannis M. and Adriaan Soetevent. "Wage and Employment in a Random Social Network with Arbitrary Degree Distribution", *American Economic Review*, 96(2), May 2006.

Jackson, Matthew O. "The Economics of Social Networks" *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society,* Chapter 1, Volume 1, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press, 2006.

Jackson, Matthew O. "A Survey of Models of Network Formation: Stability and Efficiency", In *Group Formation in Economics: Networks, Clubs, and Coalitions*. Edited by Gabrielle Demange and Myrna Wooders, Cambridge University Press, 2005.

Jackson, Matthew O. and Brian W. Rogers. "Meeting Strangers and Friends of Friends: How Random are Socially Generated Networks?" *American Economic Review*, forthcoming.

Joyner, Kara and Grace Kao. "School Racial Composition and Adolescent Racial Homophily", *Social Science Quarterly*, 81(3), 2000.

Kremer, Michael and Dan M. Levy. "Peer Effects and Alcohol Use Among College Students", NBER Working Paper 9876, July 2003.

Lyle, David S. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point", *Review of Economics and Statistics*, 89(2), May 2007.

Manski, Charles F. "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, 14(3), Summer 2000.

Montgomery, James D. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis", *American Economic Review*, 81(5), 1991.

Moody, James. "Race, School Integration, and Friendship Segregation in America", *American Journal of Sociology*, 107(3), 2001.

Newman, Mark E.J. "The structure and function of complex networks", *Society for Industrial and Applied Mathematics Review,* 45, 2003.

Pellizarri, Michele. "Do Friends and Relatives Really Help in Getting a Job?" *London School of Economics, CEP Discussion Paper* 623, 2004.

Quillian, Lincoln and Mary E. Campbell. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation", *American Sociological Review*, 68, 2003.

Rosenblat, Tanya S. and Markus M. Mobius. "Getting Closer or Drifting Apart?", *Quarterly Journal of Economics*, 119(3), August 2004.

Sacerdote, Bruce I. "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 116(2), May 2001.

Sacerdote, Bruce I. and David Marmaros. "How Do Friendships Form?", *Quarterly Journal of Economics*, 121(1), February 2006.

Soetevent, Adriaan R. "Social Interactions and Economic Outcomes", thesis, University of Groningen, 2004.

Stinebrickner, Todd R. and Ralph Stinebrickner. "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds", *Journal of Public Economics*, Vol. 90, 2006.

Ward, Bryce A. "Distance and Social Capital: Can Isolation Be Good?", mimeo, Harvard University, 2004.

Weinberg, Bruce A. "Social Interactions and Endogenous Association", mimeo, Ohio State University, 2005.

Winston, Gordon C. and David J. Zimmerman. "Peer Effects in Higher Education" NBER Working Paper 9501, February 2003.

Wasserman, Stanley and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

Zimmerman, David J. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics*, 85(1), February 2003.

Appendix: Robustness checks of the Model of Network Formation

When simulating the model of network formation we assume that the realizations u_{ij} are independent for each pair *i/j*. To check whether the results hinge on this assumption, we recalibrate the model while imposing a connection between u_{ij} and u_{ik}, and u_{jk}. We generate this connection by calculating the distance between two individuals based on a random draw of a one dimensional characteristic.[38] Each individual is randomly assigned a position on a circle. The bigger the distance between two individuals along this circle, the less likely they are forming a friendship. By construction, a small distance between *i* and *j,* and between *i* and *k* implies a small distance between *j* and *k*; the likelihood of friendship formation is correlated. We normalize the maximum distance between two individuals to one. Consequently, the distance between two individuals is distributed uniform over the interval between zero and one. Using the inverse of the standard normal CDF we transform the uniformly distributed distance measure into a standard normal distributed value u_{ij}^{Corr}. We combine this value with a randomly drawn value u_{ij}^{Random} to generate u_{ij}^{*}:[39]

$$u_{ij}^{*} = \left(u_{ij}^{Corr} + au_{ij}^{Random}\right)\frac{1}{\sqrt{a+1}}.$$

We pick *a=10* and recalibrate the model using u_{ij}^{*} instead of u_{ij}. We use the resulting parameters to construct the same counterfactuals as in the original specification of the model (see Tables 7 and 8). The results are displayed in Appendix Table 2. When comparing the results of the counterfactual experiments under uncorrelated preferences (Table 8) to the counterfactuals under correlated preferences (Appendix Table 2), we see few differences. Some of the counterfactuals

[38] This is an extreme case. Two individuals may like or dislike each other based on a vast number of different characteristics. The other extreme is the case with *n* different independent characteristics which leads less correlation in u_{ij} and is similar to the specification used in the original version.

[39] This procedure introduces a connection between u_{ij}^{*} and u_{ik}^{*}, and u_{jk}^{*}, while preserving the properties of the distribution like mean zero for each individual and normal distribution.

result in slightly different distributions of the number of friends than under the original specification. The correlated preferences lead to a bigger role for the friends of friends channel. Comparing columns (2), (4), (5), and (6) in the two tables reveals that with correlated preferences, the variance and skewness are relatively more influenced by the friends of friends channel than by the different meeting channels or preferences. The increased importance of friends of friends and a lesser role of the meeting channels diminishes the effectiveness of the special introduction of minorities to whites (column 8). In the original specification, the fraction of white friends of whites drops from 87% to 77%. With correlated preferences, the same policy results in a drop to 81%. By assuming uncorrelated preferences we might overstate the potential effect of the introducing minority students to whites. Otherwise, relaxing the non-correlation assumption does not change the conclusion that university policies have limited ability to reduce the observed segregation. The effects of the "random meeting" counterfactual, and the "affirmative action" counterfactual are similar under both correlated and uncorrelated preferences. Overall, these comparisons suggest that while allowing for correlated preferences can have some effects on our results, the qualitative conclusions do not change.

We also calibrate a specification of the model where the number of friends of friends that are met in each cycle depends on the number of current friends of an individual. The probability, $p_{frofr(i)}$, that person *i* meets each of her friends' friends is given by:

$$p_{frofr(i)} = \overline{p}_{frofr}\left(1+\frac{1}{1+\#CF_{(i)}}\right)\frac{1}{2}.$$

Where $\overline{p}_{frofr}$ is a parameter that is calibrated. The actual meeting probability $p_{frofr(i)}$ is a function of this parameter and the number of current friends of *i*, $CF_{(i)}$. The network parameters and resulting counterfactual simulations are very similar to ones obtained in the original specification.

Table 1: Characteristics of 10 Universities in Our Sample and Facebook Uptake Rates

	Rice	U Texas	Texas A&M	Baylor	Texas Tech	Texas Christian	SMU	U North Texas	UT Arlington	Texas State
Enrollment:	2,933	36,473	35,605	11,521	23,329	7,024	6,090	24,274	18,176	22,402
Female:	49%	52%	49%	58%	45%	60%	55%	55%	53%	55%
Out of State:	47%	5%	3%	16%	4%	20%	31%	3%	2%	1%
International:	3%	3%	1%	1%	1%	4%	4%	3%	5%	1%
Black:	7%	4%	2%	7%	3%	5%	5%	12%	14%	5%
Asian:	15%	17%	3%	6%	2%	2%	6%	4%	11%	2%
Caucasian:	55%	60%	82%	74%	81%	79%	75%	68%	57%	71%
Hispanic:	11%	15%	10%	8%	11%	6%	9%	10%	13%	19%
Native American:	1%	0%	1%	1%	1%	1%	1%	1%	1%	1%
Undergrads in Facebook Data:	2,354	14,728	15,797	7,008	7,219	3,678	3,496	4,474	1,442	4908
Facebook Uptake Rate:	80%	40%	44%	61%	31%	52%	57%	18%	8%	22%

Source: Princeton Review (2005). The universities are listed in the order that they joined the Facebook. The private universities are Rice, Baylor, Texas Christian and Southern Methodist. Facebook uptake rates includes undergraduate students registered for facebook.com as of January 17, 2005.

Table 2: Network Composition and Chartacteristics

	Rice	U Texas	Texas A&M	SMU	Baylor	Texas Tech	Texas Christian	U North Texas	UT Arlington	Texas State
Composition										
Number of Students	1300	8467	9299	2223	4295	4648	2342	2607	820	2922
Fraction female	0.50	0.56	0.55	0.59	0.60	0.53	0.64	0.57	0.47	0.58
Fraction White or Hispanic	0.82	0.85	0.96	0.94	0.91	0.97	0.95	0.92	0.82	0.96
Fraction Black	0.05	0.02	0.02	0.04	0.06	0.02	0.04	0.06	0.12	0.03
Fraction Asian	0.13	0.13	0.02	0.03	0.03	0.01	0.01	0.02	0.06	0.01
Fraction liberal	0.32	0.23	0.06	0.11	0.08	0.08	0.12	0.18	0.16	0.14
Fraction conservative	0.15	0.23	0.54	0.42	0.47	0.50	0.43	0.27	0.27	0.31
Characteristics										
Average number friends	50.8	39.5	41.1	62.9	59.8	40.5	49.8	23.8	17.2	25.6
Variance of number of friends	31.9	36.5	38.4	48.3	50.8	35.6	36.0	23.9	17.7	23.8
Skewness of number of friends	1.06	2.01	2.06	1.75	1.74	1.50	1.11	2.28	1.52	1.69
Total Cluster coefficient	0.24	0.20	0.17	0.23	0.19	0.21	0.23	0.21	0.27	0.23
Avg. Individual cluster coefficient	0.30	0.22	0.19	0.27	0.21	0.23	0.25	0.22	0.25	0.23
Degree correlation	0.22	0.57	0.57	0.49	0.58	0.57	0.54	0.35	0.53	0.55
Degree-cluster correlation	-0.47	-0.04	-0.10	-0.17	-0.14	-0.08	-0.09	0.06	0.16	0.02

Note: This table includes undergraduates in our facebook.com sample for whom we could identify race based upon the picture. Degree, degree-correlation, total cluster coefficient, and individual cluster coefficient are defined in section II.B. The degree-cluster correlation is the correlation between students' number of friends (degree) and individual cluster coefficient.

Table 3: Segmentation by Race

	Rice	U Texas	Texas A&M	SMU	Baylor	Texas Tech	Texas Christian	U North Texas	UT Arlington	Texas State
Pair of:	**Fraction of pairs who are friends (in %)**									
White/Hisp & White/Hisp	4.05	0.52	0.45	2.97	1.53	0.89	2.20	0.95	2.40	0.89
White/Hisp & Asian	3.08	0.20	0.33	1.72	0.60	0.46	1.17	0.70	0.67	0.74
White/Hisp & Black	3.40	0.26	0.34	1.50	0.57	0.61	1.38	0.60	1.16	0.66
Asian & Asian	9.43	1.93	3.28	17.69	5.90	3.36	5.21	3.28	3.35	1.56
Asian & Black	3.61	0.25	0.45	2.43	0.72	0.70	1.65	0.63	0.77	0.88
Black & Black	20.04	6.13	7.31	19.59	8.34	6.41	11.90	4.61	12.01	5.55
Any two students	3.92	0.47	0.44	2.83	1.39	0.87	2.13	0.92	2.10	0.88
Pair of:	**Relative probability of friendship**									
White/Hisp & White/Hisp	1.03	1.12	1.01	1.05	1.10	1.02	1.03	1.04	1.14	1.01
White/Hisp & Asian	0.79	0.42	0.74	0.61	0.43	0.52	0.55	0.77	0.32	0.84
White/Hisp & Black	0.87	0.56	0.77	0.53	0.41	0.70	0.65	0.66	0.55	0.75
Asian & Asian	2.41	4.13	7.42	6.24	4.23	3.85	2.45	3.58	1.59	1.78
Asian & Black	0.92	0.54	1.01	0.86	0.52	0.80	0.77	0.69	0.36	1.00
Black & Black	5.12	13.13	16.54	6.92	5.99	7.35	5.59	5.03	5.71	6.33
Any two students	1	1	1	1	1	1	1	1	1	1
Fraction of Students White/Hisp	0.82	0.85	0.96	0.94	0.91	0.97	0.95	0.92	0.82	0.96
Fraction Friends of Wh/H who are Wh/H	0.85	0.93	0.97	0.96	0.96	0.98	0.97	0.94	0.92	0.97
Fraction of Students Asian	0.13	0.13	0.02	0.03	0.03	0.01	0.01	0.02	0.06	0.01
Fraction Friends of Asians who are Asian	0.30	0.58	0.16	0.22	0.25	0.07	0.05	0.10	0.23	0.02
Fraction of Students Black	0.05	0.02	0.02	0.04	0.06	0.02	0.04	0.06	0.12	0.03
Fraction Friends of Blacks who are Black	0.25	0.38	0.27	0.32	0.47	0.17	0.25	0.33	0.58	0.18

Note: This table includes undergraduates in our facebook.com sample for whom we could identify race based upon the picture. Students were classified as either White/Hispanic, Black, Asian, or Don't Know, as described in Section II.B. The fraction of pairs of students of race X and Y who are friends is the fraction of all possible pairs of students of race X and Y who report being friends (reported in percentage points). The relative probability of friendship is defined in section II.B.

Table 4: Segmentation by Major, Cohort and Political Orientation

	Rice	U Texas	Texas A&M	SMU	Baylor	Texas Tech	Texas Christian	U North Texas	UT Arlington	Texas State
Segmentation by Major										
Fraction of Friends in Same Major if friendships were formed randomly	0.04	0.02	0.02	0.01	0.02	0.03	0.01	0.01	0.05	0.01
Actual Fraction of Friends in Same Major	0.08	0.08	0.07	0.08	0.06	0.06	0.07	0.08	0.10	0.08
Segmentation by Cohort										
Pair of:	**Relative probability of friendship**									
Freshman & Freshman	2.14	2.24	2.10	2.10	2.10	2.01	1.95	1.85	1.72	2.07
Freshman & Sophomore	0.64	0.74	0.72	0.64	0.60	0.82	0.74	0.84	1.00	0.79
Freshman & Junior	0.46	0.40	0.45	0.38	0.33	0.52	0.45	0.62	0.73	0.46
Freshman & Senior	0.35	0.25	0.31	0.20	0.18	0.43	0.25	0.58	0.61	0.31
Sophomore & Sophomore	2.18	2.28	2.04	2.42	2.62	1.80	2.19	1.74	1.29	2.01
Junior & Junior	2.17	2.13	2.14	2.21	2.29	1.46	2.17	1.55	1.27	1.77
Senior & Senior	1.80	2.05	2.43	2.08	2.06	1.71	1.92	2.38	1.95	1.93
Segmentation by Political Orientation										
Pair of:	**Relative probability of friendship**									
Liberal & Liberal	1.22	1.06	1.28	1.00	1.13	1.07	1.09	1.18	1.24	1.05
Liberal & Conservative	0.86	0.75	0.69	0.66	0.59	0.70	0.85	0.76	0.79	0.81
Conservative & Conservative	1.35	2.17	1.28	1.36	1.41	1.44	1.30	1.45	1.84	1.53

See Notes of Table 3.

Table 5: Selection into Facebook at Texas A&M - Means

	Students In Facebook	Overall Student Population
GPA	2.95	2.93
SAT	1168	1152
High School %ile Class Rank	86.5	86.0
High School Class Size	437	416
Texas Resident	97.4%	97.4%
Female	55.2%	50.6%
In a Greek	14.3%	11.6%
In Corps	1.7%	1.8%
Lives in a dorm	41.1%	33.7%
Athlete	2.5%	2.5%
Freshman	27%	22%
Sophomore	27%	22%
Junior	26%	26%
Senior	20%	29%
White	81.8%	80.5%
Hispanic	11.4%	12.0%
Asian	4.0%	3.8%
Black	2.3%	2.9%
Native American	0.4%	0.5%
Father College Degree	61%	58%
Mother College Degree	54%	51%
Household Income < $40,000	14%	17%
Household Income $40,000-$80,000	33%	35%
Household Income > $80,000	53%	48%
N	6754	17288

The sample is all undergraduate students enrolled in Spring 2005 with complete data for gender, high school rank, SAT, parents' education, household income, financial need, Greek and Corps status. 54% of registered students have complete data for these variables. Students with complete data are relatively similar to those without complete data in GPA (2.93 vs. 2.87), SAT (1152 vs. 1147), father college (58% vs. 56%), mother college (51% vs. 49%), but differ in high school rank (86% vs 80%), Greek (12% vs 15%), and parental income less than $40K (17% vs 8%).

Table 6: Factors Predicting the Probability that Two Students are Friends

Dependent Variable =1 if students *i* and *j* are friends and =0 otherwise
Mean of Dependent Variable (baseline rate): 0.0034

Relationship Between Student *i* and *j*

	Race	High School, Age	Family	Dorm, Academic	Ability	Activities	All	Common Friends?
Constant	0.0036 **	0.0039 **	0.0023 **	0.0028 **	0.0045 **	0.0032 **	0.0032 **	-0.0005 **
Both Black	0.0551 **						0.0537 **	0.0153 **
Both Asian	0.0122 **						0.0121 **	0.0074 **
Both Hispanic	0.0017 **						0.0021 **	0.0016 **
Both Native Am	-0.0036 **						-0.0038 **	-0.0008
White-Hispanic	-0.0010 **						-0.0003 **	0.0001
White-Asian	-0.0012 **						-0.0007 **	0.0000
White-Black	-0.0012 **						-0.0006 **	-0.0004
White-NatAm	-0.0010						-0.0008	0.0005
Hispanic-Asian	-0.0009 **						0.0000	0.0003
Hispanic-Black	0.0000						0.0005	0.0000
Hispanic-NatAm	-0.0012 *						-0.0005	0.0009
Asian-Black	-0.0002						0.0005	-0.0002
Asian-NatAm	-0.0020 **						-0.0014 **	0.0003
Black-NatAm	-0.0018						-0.0015	0.0001
Same High School		0.1864 **					0.1859 **	0.1379 **
Same Year in College		0.0010 **					0.0011 **	0.0012 **
Same Gender		0.0006 **					0.0000	-0.0005 **
Difference b/t Yrs in College (Yrs)		-0.0013 **					-0.0011 **	0.0001 **
Both from High Income Households (>$60K)			0.0005 **				0.0002	-0.0003 **
Both from Low Income Households (<$60K)			0.0003 **				0.0003 **	0.0003 **
2 College Parents - 2 College Parents			0.0013 **				0.0009 **	-0.0013 **
2 College Parents - 1 College Parent			0.0004 **				0.0003	-0.0008 **
1 College Parent - 1 College Parent			0.0002				0.0001	-0.0004 **
2 College Parents - 0 College Parents			-0.0001				0.0000	-0.0006 **
1 College Parent - 0 College Parents			-0.0001				-0.0001	-0.0003 **
Students Both Liberal			0.0025 **				0.0021 **	0.0017 **
Students Both Conservative			0.0023 **				0.0019 **	-0.0012 **
Students One Liberal One Conservative			-0.0002				-0.0001	-0.0003
Same Dorm				0.0426 **			0.0406 **	0.0214 **
Same Major				0.0038 **			0.0030 **	0.0024 **
Same College on Campus				0.0018 **			0.0016 **	0.0004 **
Difference in SAT scores (absolute points in 100s)					-0.0004 **		-0.0003 **	0.0000
Difference in GPA Quintile (0-4 absolute quintiles)					-0.0003 **		-0.0002 **	-0.0001 **
Both are Athletes						0.0646 **	0.0633 **	0.0110 **
Both in Corps of Cadets						0.0531 **	0.0421 **	0.0218 **
Both are Greek						0.0189 **	0.0183 **	-0.0082 **
One is Greek						-0.0003 **	-0.0004 **	-0.0022 **
One is Athlete						-0.0003	-0.0003	-0.0015 **
One in Corps of Cadets						-0.0005	-0.0006	-0.0005
Number of Common Friends								0.0299 **
R^2	0.0006	0.0293	0.0004	0.0033	0.0001	0.0032	0.0362	0.2457

* significant at 5% ** significant at 1%

Obervations are all pairwise combinations of students in Texas A&M Facebook with complete data on covariates (29,787,621=N*(N-1)/2 observations where N=7719). Linear probability model estimated via least squares. Bootstrap confidence intervals are constructed by sampling with replacement over individual students to obtain 7719 students and forming all pairwise combinations of those students as the bootstrap sample. We construct 200 bootstrap samples. Table only reports coefficient estimates and significance levels to conserve space but confidence intervals are available upon request. Excluded category for race is white-white and for political orientation is no reported orientation.

Table 7: Parameters of the Model Under the Calibration and the Counterfactual Experiments

			Counterfactuals					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample of 1930 Students	Full Model Simulation	Completely Random Friends	Full Model without friends of friends	Random Meeting	No Preferences	Affirmative Action, double hispanics	Introduction to students of different race
Number of Cycles of Meeting Friends of Friends	-	8	0	0	8	8	8	8
c_{init} - Avg # Randomly Met	-	6.15	6.41	14.58	25.56	6.16	5.70	4.98
Probability of meeting friend of friend	-	0.54	0	0	0.54	0.54	0.54	0.54
c_{coll} - Avg # Met Same College	-	4.60	0	10.90	0	4.60	4.26	3.73
Probability of meeting student in same Year	-	0.02	0	0.05	0	0.02	0.02	0.02
Probability of meeting student in same dorm	-	0.35	0	0.83	0	0.35	0.32	0.28
$\beta_{Constant}$	-	-1.72	0	-1.72	-1.72	-1.57	-1.72	-1.72
β_{WW} (Whites)	-	0.07	0	0.07	0.07	0	0.07	0.07
β_{BB} (Blacks)	-	2.10	0	2.10	2.10	0	2.10	2.10
β_{HH} (Hispanics)	-	0.40	0	0.40	0.40	0	0.40	0.40
β_{AA} (Asians)	-	0.85	0	0.85	0.85	0	0.85	0.85
β_{skill} (High SAT)	-	0.10	0	0.10	0.10	0	0.10	0.10
β_{ParEdu}	-	0.09	0	0.09	0.09	0	0.09	0.09
$\beta_{Conservative}$	-	0.12	0	0.12	0.12	0	0.12	0.12

Notes:

The data used are a random sample of 1930 of the 7719 students described in the note to Table 8.

(2) are the parameters of the full model calibration that fit the simulated moments to the moments of the actual network.

(3) Students meet with the same probability independent of school environment, they do not have preferences for characteristics, and they do not meet friends of friends.

(4) Students meet with probabilities that vary with the school environment, they have preferences for characteristics, but they do not meet friends of friends.

(5) Students meet with the same probability independent of school environment, they have preferences for characteristics, and they meet friends of friends.

(6) Students meet with probabilities that vary with school environment, they do not have preferences for characteristics, and they meet friends of friends.

(7) Double the number of Hispanic students (with parameters of full model but meeting probabilities scaled down to generate the same average number of friends).

(8) Add an extra meeting round where each white meets 1% of minority students and each minority student meet 1% of white students (with parameters of full model but meeting probabilities scaled down to generate same average number of friends).

Table 8: Simulation Results and Counterfactual Experiments

			Counterfactuals					
	(1)	**(2)**	(3)	(4)	(5)	(6)	(7)	(8)
Moments Entering Calibration	Sample of 1930 Students	**Full Model Simulation**	Completely Random Friends	Full Model without friends of friends	Random Meeting	No Preferences	Affirmative Action, double hispanics	Introduction to students of different race
Average # of Friends	6.42	**6.42**	6.41	6.42	6.41	6.42	6.41	6.41
Variance of # of Friends	6.44	**6.27**	2.52	2.96	5.56	5.77	6.40	6.14
Skewness of # of Friends	1.82	**1.82**	0.39	0.69	1.58	1.56	1.88	1.79
Cluster Coefficient	0.15	**0.16**	0.00	0.01	0.17	0.16	0.16	0.17
Fraction from Same Year	0.44	**0.44**	0.25	0.59	0.25	0.44	0.45	0.39
Fraction from Same College	0.22	**0.22**	0.13	0.31	0.13	0.21	0.21	0.20
Fraction from Same Dorm	0.08	**0.07**	0.01	0.14	0.01	0.07	0.08	0.06
Fraction White Friends of Whites	0.87	**0.85**	0.82	0.85	0.85	0.82	0.76	0.77
Fraction Hispanic Friends of Hispanics	0.21	**0.22**	0.12	0.23	0.22	0.12	0.42	0.21
Fraction Asian Friends of Asians	0.15	**0.14**	0.04	0.14	0.14	0.03	0.12	0.14
Fraction Black Friends of Blacks	0.32	**0.33**	0.02	0.22	0.28	0.02	0.28	0.31
Fraction Hi SAT Score Friends of Hi SAT	0.49	**0.49**	0.39	0.47	0.47	0.41	0.47	0.48
Fraction Friends of Same Parental Education	0.53	**0.53**	0.44	0.50	0.52	0.45	0.50	0.51
Fraction Conservative Friends of Conservative	0.62	**0.62**	0.52	0.59	0.61	0.53	0.60	0.60
Other Moments Not Entering Calibration								
Fraction White Friends of Whites if same Year	0.87	**0.86**	0.82	0.84	0.85	0.83	0.74	0.80
Fraction Hispanic Friends of Hispanics if same Year	0.23	**0.25**	0.11	0.25	0.25	0.15	0.48	0.19
Fraction White Friends of Whites if same College	0.87	**0.87**	0.82	0.84	0.85	0.84	0.77	0.80
Fraction Hispanic Friends of Hispanics if same College	0.22	**0.26**	0.13	0.24	0.16	0.12	0.50	0.15
Fraction White Friends of Whites if same Dorm	0.84	**0.86**	0.78	0.84	0.82	0.83	0.78	0.83
Fraction Hispanic Friends of Hispanics if same Dorm	0.17	**0.28**	0.00	0.22	0.00	0.14	0.47	0.16
Average Individual Cluster Coefficient	0.12	**0.16**	0.00	0.01	0.16	0.16	0.15	0.15
Degree correlation	0.49	**0.61**	-0.02	0.23	0.63	0.62	0.70	0.64
Degree-cluster correlation	0.24	**0.15**	0.05	0.17	0.21	0.16	0.10	0.28

See notes in Table 7 for a description of each counterfactual.

Table 9: Associations Between Student Outcomes and Peer Characteristics

Dependent Variable:	Own GPA		Drinker	Volunteer		Religious		Political	
				Excl. Same Orgs	Incl. Same Orgs	Excl. Same Orgs	Incl. Same Orgs	Excl. Same Orgs	Incl. Same Orgs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OWN CHARACTERISTICS									
SAT in 100s of Points	0.096	0.093	-0.006	0.002	0.002	0.004	0.003	-0.005	-0.006
	[0.007]**	[0.007]**	[0.004]	[0.004]	[0.004]	[0.004]	[0.003]	[0.023]	[0.023]
High School Percentile (0-100)	0.015	0.015	-0.001	0.001	0.001	0	-0.001	0.002	0.002
	[0.001]**	[0.001]**	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.003]
Female	0.135	0.117	-0.032	0.061	0.055	0.016	0.015	-0.464	-0.465
	[0.017]**	[0.016]**	[0.010]**	[0.010]**	[0.010]**	[0.009]	[0.009]	[0.058]**	[0.058]**
Parent HH Income $40-80K	-0.039	-0.037	0.016	0.007	0.007	-0.009	-0.009	0.003	0.004
	[0.025]	[0.025]	[0.014]	[0.015]	[0.015]	[0.013]	[0.013]	[0.086]	[0.086]
Parent HH Income > $80K	0.009	0.008	0.031	0.006	0.002	-0.016	-0.014	0.004	0.003
	[0.026]	[0.026]	[0.014]*	[0.015]	[0.015]	[0.014]	[0.013]	[0.088]	[0.088]
Father College Grad	0.068	0.060	-0.008	0.003	0.001	0.024	0.019	0.097	0.098
	[0.018]**	[0.018]**	[0.010]	[0.011]	[0.010]	[0.010]*	[0.009]*	[0.061]	[0.061]
Mother College Grad	0.015	0.015	0.002	0.009	0.01	0.006	0.006	-0.05	-0.049
	[0.016]	[0.016]	[0.009]	[0.010]	[0.010]	[0.009]	[0.009]	[0.057]	[0.057]
High School Pct Economically Disadvantaged	-0.003	-0.003	0.000	0.000	0.000	0.000	0.000	-0.002	-0.002
	[0.001]**	[0.001]**	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.002]
High School Pass Rate of Standardized TAAS Test	0.006	0.005	0.000	0.001	0.000	0.001	0.001	-0.009	-0.009
	[0.001]**	[0.001]**	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.005]	[0.005]
FRIEND CHARACTERISTICS									
Predetermined									
Avg of Friends' SAT in 100s of Points	0.030	-0.032	-0.014	0.006	0.009	-0.004	-0.016	-0.088	-0.089
	[0.016]	[0.017]	[0.009]	[0.010]	[0.010]	[0.009]	[0.009]	[0.060]	[0.059]
Avg of Friends' High School Percentile	0.002	-0.005	-0.002	0.000	-0.001	0.002	0.001	0.005	0.006
	[0.002]	[0.002]**	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.007]	[0.007]
Fraction of Friends with College Educated Father	0.160	0.106	-0.019	0.059	0.028	0.099	0.061	0.034	0.030
	[0.058]**	[0.057]	[0.032]	[0.040]	[0.039]	[0.036]**	[0.035]	[0.231]	[0.231]
Fraction of Friends with College Educated Mother	0.029	-0.007	0.032	0.049	0.038	0.036	0.046	0.498	0.494
	[0.057]	[0.056]	[0.032]	[0.039]	[0.039]	[0.036]	[0.034]	[0.228]*	[0.228]*
Contemporaneous									
Avg of Friends' GPA		0.461							
		[0.044]**							
Fraction of Friends who 'Drink'		-0.123	0.386						
		[0.075]	[0.042]**						
Fraction of Friends in Volunteer Groups				-0.036	0.564				
				[0.045]	[0.042]**				
Fraction of Friends in Religious Groups						0.497	0.781		
						[0.033]**	[0.028]**		
Fraction of Friends in Political Groups								0.200	0.206
								[0.418]	[0.412]
Constant	-0.612	-0.456	0.573	-0.199	-0.222	-0.306	-0.081	3.789	3.76
	[0.230]**	[0.230]*	[0.140]**	[0.153]	[0.150]	[0.140]*	[0.133]	[0.887]**	[0.886]**
Observations	5138	5138	5138	4655	4661	4655	4661	4655	4661
R-squared	0.31	0.33	0.04	0.03	0.07	0.07	0.16	0.04	0.04

* significant at 5%   ** significant at 1%

Notes: Models estimated via least squares with Huber-White robust standard errors reported in brackets. The dependent variable in the last 4 sets of models is a dummy variable for whether the student is a drinker or a member of a volunteer, religious, or politica organization. All regressions include dummies for race, year in college, academic college on campus (e.g. Liberal Arts or Engineering), the number of members in the student's family household, and the Spectral Segregation Index (SSI) for each race. The sample includes all students among 7719 (used in the model in sections II and III) who have friends among the 7719 to compute average friend characteristics. "Excl. Same Orgs" and "Incl. Same Orgs" means that the calculation of the fraction of friends in the type of organization either *excludes* or *includes* friends who are in the same organization.

Appendix Table 1: Testing Accuracy of Using Pictures to Classify Race

Administrative Data	**Classification by Picture**				
	White/Hispanic	Black	Asian	Other/Don't Know	% Mis-classified
White/Hispanic	4,174	3	4	73	0.2%
Black	3	92	0	4	3.0%
Asian	14	0	90	36	10.0%
Native American	18	0	0	0	--

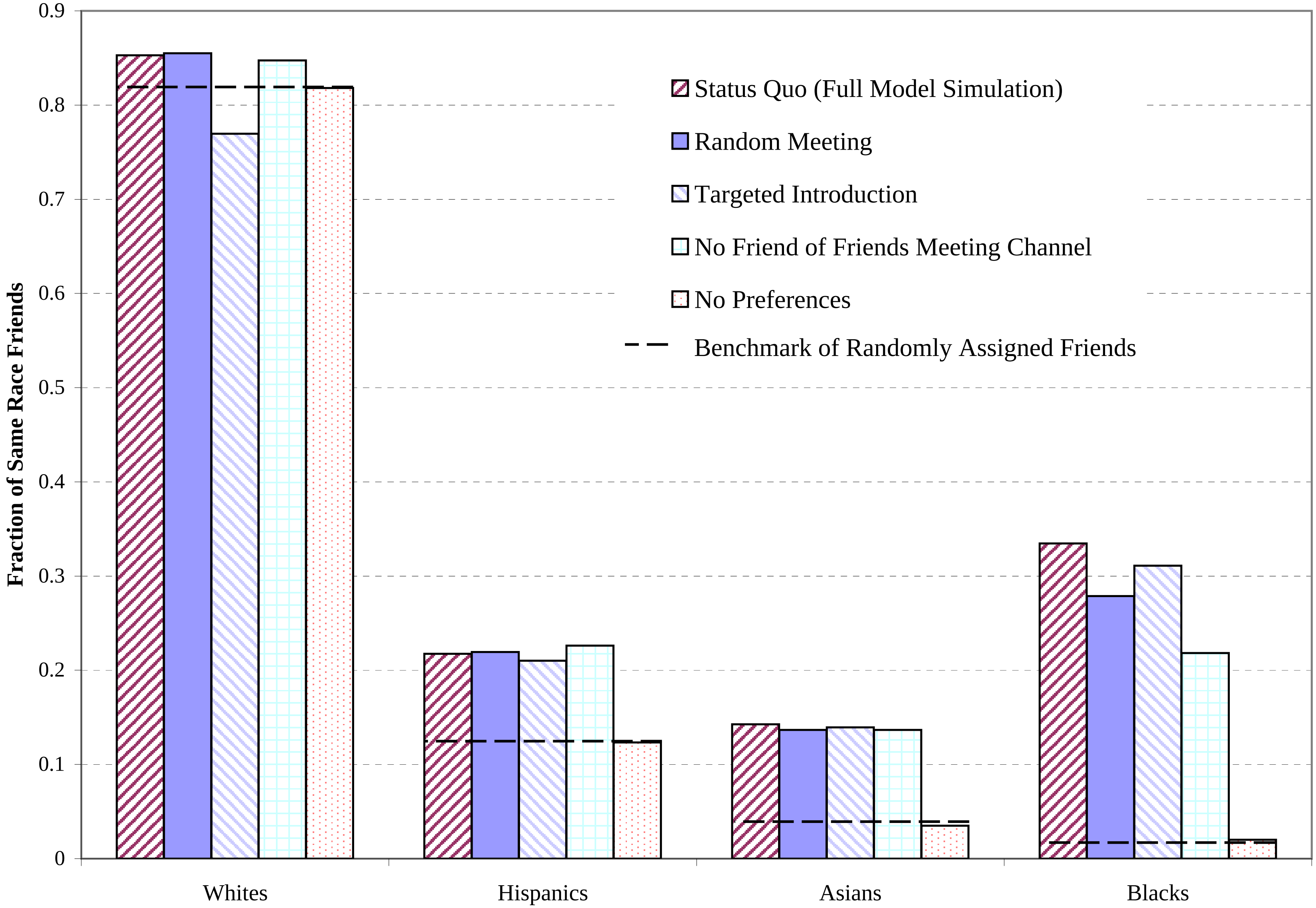
Notes: This table compares the actual race from administrative records to the race classification made by research assistants based upon pictures. The sample is all Texas A&M students meeting two criteria: (1) in Facebook with a picture in which the two research assistants agreed on their race classification and (2) administrative records contain complete demographic data. In the picture classification data, "Other" includes those classified as 'don't know but not black' and 'don't know but maybe black'. The "% Mis-classified" is the percent of students in each category of Administrative race that are improperly classified in a specific category (we do not include students that the research assistant classified as Don't Know).

Appendix Table 2: Simulation Results and Counterfactual Experiments for Correlated Preferences

			Counterfactuals					
	(1)	(2)	(3)	(5)	(4)	(6)	(7)	(8)
	Sample of 1930 Students	Full Model Simulation	Completely Random Friends	Full Model without friends of friends	Random Meeting	No Preferences	Affirmative Action, double hispanics	Introduction to students of different race
Moments Entering Calibration								
Average # of Friends	6.42	6.41	6.42	6.42	6.42	6.41	6.42	6.41
Variance of # of Friends	6.44	6.25	2.53	2.84	5.83	5.74	6.59	6.16
Skewness of # of Friends	1.82	1.83	0.39	0.60	1.77	1.62	2.01	1.77
Cluster Coefficient	0.15	0.15	0.00	0.01	0.15	0.15	0.15	0.15
Fraction from Year	0.44	0.43	0.25	0.59	0.25	0.44	0.45	0.41
Fraction form College	0.22	0.22	0.13	0.31	0.14	0.22	0.22	0.21
Fraction from sam dorm	0.08	0.08	0.01	0.11	0.01	0.07	0.08	0.07
Fraction White Friends of Whites	0.87	0.86	0.82	0.85	0.86	0.82	0.76	0.81
Fraction Hispanic Friends of Hispanics	0.21	0.21	0.12	0.22	0.21	0.12	0.40	0.21
Fraction Asian Friends of Asians	0.15	0.15	0.04	0.15	0.16	0.04	0.13	0.16
Fraction Black Friends of Blacks	0.32	0.33	0.02	0.22	0.24	0.02	0.26	0.32
Fraction Hi SAT Score Friends of Hi SAT	0.49	0.49	0.39	0.47	0.48	0.41	0.48	0.49
Fraction Friends of Same Parental Education	0.53	0.53	0.44	0.50	0.52	0.44	0.50	0.52
Fraction Conservative Friends of Conservative	0.62	0.63	0.52	0.59	0.62	0.53	0.61	0.61

Notes: This table includes results of model calibration under the assumption of **correlated unobserved preferences**, as described in the Appendix. The calibration is based on the same sample as the original model (see results in Table 7), and counterfactuals are equivalent to those performed for the original model (see Table 8).
The parameters of the fully calibrated model are:
$CYCLE = 11$, $c_{init} = 10.68$, $c_{COLL} = 7.52$, $p_{YEAR} = .03$, $p_{DROM} = .59$, $p_{FRoFR} = .63$,
$\beta_0 = -1.92$, $\beta_{WW} = .07$, $\beta_{BB} = 1.75$, $\beta_{HH} = .36$, $\beta_{AA} = .86$, $\beta_{PAR\_edu} = .09$, $\beta_{CONS} = .12$, $\beta_{SKILL} = .10$

Figure 1: Counterfactual Experiments: Fraction of Same Race Friends



See Notes to Table 7 for a description of each counterfactual experiment.