Identifying the roles of race-based choice and chance in high school friendship network formation

Sergio Currarini[a], Matthew O. Jackson[b,c,1], and Paolo Pin[d]

[a]Dipartimento di Scienze Economiche, Università Ca′ Foscari di Venezia, and School for Advanced Studies in Venice, 30123 Venice, Italy; [b]Department of Economics, Stanford University, Stanford, CA 94305; [c]Santa Fe Institute, Santa Fe, NM 87501; and [d]Dipartimento di Economia Politica, Università degli Studi di Siena, 53100 Siena, Italy

Edited* by Christos Papadimitriou, University of California, Berkeley, CA, and approved January 13, 2010 (received for review October 12, 2009)

Homophily, the tendency of people to associate with others similar to themselves, is observed in many social networks, ranging from friendships to marriages to business relationships, and is based on a variety of characteristics, including race, age, gender, religion, and education. We present a technique for distinguishing two primary sources of homophily: biases in the preferences of individuals over the types of their friends and biases in the chances that people meet individuals of other types. We use this technique to analyze racial patterns in friendship networks in a set of American high schools from the Add Health dataset. Biases in preferences and biases in meeting rates are both highly significant in these data, and both types of biases differ significantly across races. Asians and Blacks are biased toward interacting with their own race at rates >7 times higher than Whites, whereas Hispanics exhibit an intermediate bias in meeting opportunities. Asians exhibit the least preference bias, valuing friendships with other types 90% as much as friendships with Asians, whereas Blacks and Hispanics value friendships with other types 55% and 65% as much as same-type friendships, respectively, and Whites fall in between, valuing other-type friendships 75% as much as friendships with Whites. Meetings are significantly more biased in large schools (>1,000 students) than in small schools (<1,000 students), and biases in preferences exhibit some significant variation with the median household income levels in the counties surrounding the schools.

friendships | high schools | homophily | segregation | social networks

Friendship networks from a sample of American high schools in the Add Health national survey[†] exhibit a strong pattern: students tend to form friendships with other students of their same ethnic group at rates that are substantially higher than their population shares (Fig. 1) (1–4). This feature, referred to as "homophily" in the sociological literature (5), is prevalent across many applications and can have important implications for behaviors (6–9). The widespread presence of homophily indicates that friendship formation differs substantially from a process of uniformly random assortment. Two key sources of homophily are (*i*) biases in individual preferences for which relationships they form and (*ii*) biases in the rates at which individuals meet each other. It is important to identify whether homophily is primarily due to just one of these biases or to both because, for instance, this can shape policies aimed at producing more integrated high schools. In this article, we present a technique for identifying these two biases, we apply this technique to the Add Health dataset, and we estimate how preference and meeting biases differ across races.

Although there is substantial evidence that race is a salient feature in how people view each other (10), such evidence does not sort out the sources of homophily, other than indicating that student preferences could be a factor. Without detailed and reliable data on the mechanics of which students meet which others, these questions are not answered by a direct analysis of friendship data. Moreover, surveys of students asking them about their racial attitudes may not reliably reflect the choices that they make. To this end, we use a technique that is well established in economics for estimating consumers' preferences: revealed preference theory (11). One infers preferences of the individual by careful observation of the choices that they make based on the opportunities that they have. We adapt these techniques for the analysis of social behavior and friendship formation. Here we infer students' preferences by observing how the number of friendships they have changes with the racial composition of their school. We employ the friendship formation model (3), here extended to allow for different biases across race in both preferences and opportunities. Using a parameterized version of this model, we are able to distinguish between the two primary sources of homophily (that is, preference bias and meeting bias), and to measure their relative magnitudes and how these differ across races. The results illustrate these techniques and the factors leading to homophily and racial segregation patterns in friendships.

Results and Discussion

There are two patterns of homophily in the Add Health data that are important to understand (Fig. 1). First, not only is there substantial homophily, but it follows a nonlinear and nonmonotone trend with respect to group size, with low levels of homophily for groups that form very large or small fractions of a school and higher levels of homophily for groups that form an intermediate sized fraction of a school. Second, homophily patterns differ significantly across races and by school size (see *SI Text*, Fig. S1, and Tables S1–S3 for a statistical analysis).

The model of friendship formation developed in ref. 3 showed how preference biases and meeting biases lead to different patterns in the numbers of friendships that people form and the resulting homophily. Thus, by taking advantage of those differences in patterns, one can identify preference and meeting biases. Here we enrich and extend the model in such a way to be able to identify race-by-race differences in these biases. Students enter the system and randomly meet friends, leaving the process when expected gains from new meetings are outweighed by the cost of searching. One key element is that students have preferences over the racial mix of their friends. Students can value a friendship with someone of their own race differently from a friendship with someone of another race. The second element is that students may end up meeting other students of their own race at a rate that is higher than what would occur if they were meeting other searching students uniformly at random. This bias may stem from the various ways in which the meeting process can depart from a uniform random process, including self-segregation through racially homogeneous activities, as well as meeting friends of friends, etc.

Author contributions: S.C., M.O.J., and P.P. designed research, performed research, contributed new analytic tools, analyzed data, and wrote the paper, with all authors contributing equally.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: jacksonm@stanford.edu.

[†]The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7–12 in the United States during the 1994–1995 school year. Data files are available from Add Health, Carolina Population Center (addhealth@unc.edu).

This article contains supporting information online at www.pnas.org/cgi/content/full/0911793107/DCSupplemental.

ECONOMIC SCIENCES

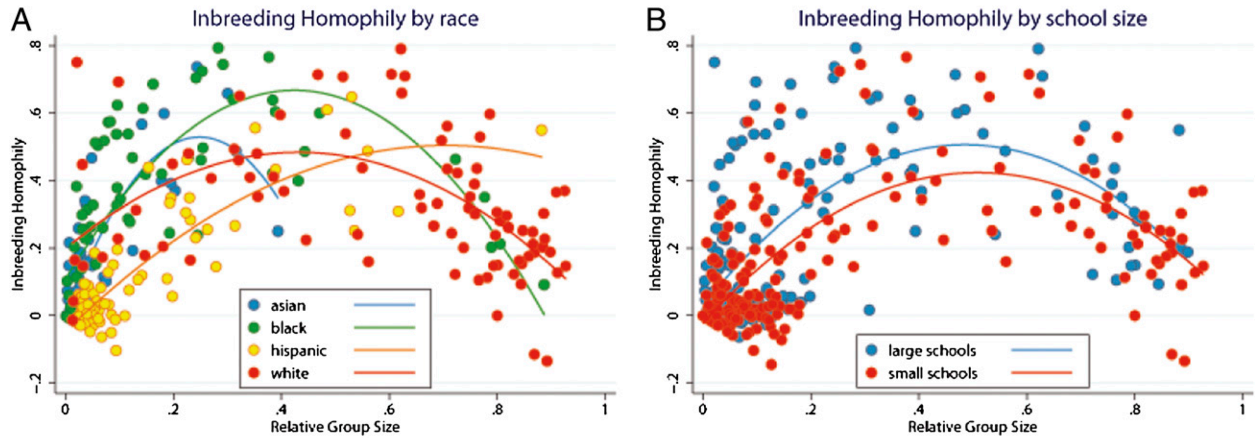

Fig. 1. Homophily as a function of the fraction of a school's population that a group comprises, differentiating by race (*A*) and school size (*B*). The homophily index, due to Coleman (12), is a normalized measure of the difference between the observed racial mix of friendships and the expected mix if friendships were formed uniformly at random. An index of 0 indicates that students in a given group (each datum is a racial group within 1 of 84 schools) have friendships distributed according to the racial mix of the society, whereas an index of 1 indicates that the students only form friendships with other students of their own race. Letting w_i be the fraction of race i in a given school, and q_i be the fraction of their friendships on average that are of own-type, the index is $(q_i - w_i)/(1 - w_i)$.

Details about the model and the way in which we estimate it from the data are found in *Material and Methods* and *SI Text*.

We find that all racial groups exhibit significant biases in both their preferences over friendships and in the rate at which they meet students of their own race. Moreover, there are significant differences across races in the relative biases.

Estimated biases in preferences range from one extreme where Blacks value friendships with students of other races 55% as much as those with other Blacks, to the other extreme where Asians value friendships with students of other races 90% as much as those with other Asians; Hispanics and Whites fall in between at 65% and 75%, respectively.

Estimated biases in meetings range from Whites meeting students without any bias, to Asians and Blacks exhibiting meeting biases of 7–7.5, and Hispanics at an intermediate rate of 2.5. To interpret the meeting biases: a meeting bias of 7 is such that >90% of the people that Asians meet are other Asians in a case where Asians comprise 50% of the population. The meeting bias of 2.5 is such that >70% of the people that Hispanics meet are other Hispanics, in a case where Hispanics comprise 50% of the population.

These results suggest that the differences across ethnic groups in both the homophily patterns and in the total number of friends (see *SI Text* for the statistical analysis of these differences) are explained by differences in both types of biases. For instance, the estimates suggest that Blacks' homophily stems from both significant meeting and preference biases, whereas Whites' homophily is more driven by preference bias.

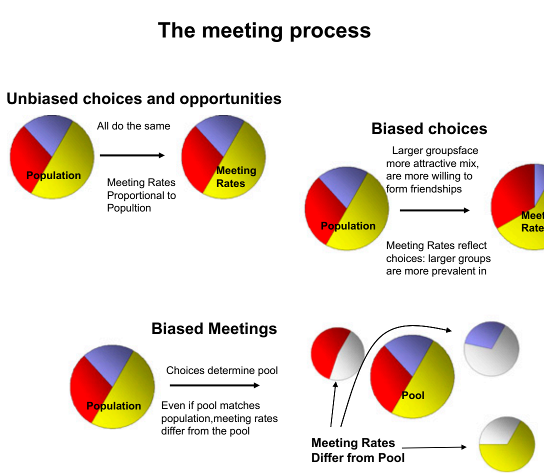

Fig. 2. The meeting process.

An additional relevant issue is the potential influence of school size on observed behavior. Hypothetically, it could be that differences across races are driven by general differences in behavior in large versus small schools, which could correlate with the racial makeup of a school and therefore potentially drive our results. We therefore control for school size, splitting the sample in "large" and "small" schools (as described in more detail below). Although controlling for school size does not affect the conclusions discussed above, it is interesting to note that larger schools exhibit significantly higher biases in the rates at which students meet students of their own race. This is consistent with more opportunities to self-segregate in large schools, where academic tracking and the availability of specialized clubs, athletics, and extracurricular activities, and other mechanisms, could bias meetings (13–19). To the extent that more integrated friendship patterns are a goal for policy, the higher bias in meetings that is observed with larger schools provides some support for the opinion expressed in some of the recent sociology literature that smaller schools offer some advantages.[‡]

We also perform an analogous control for the income level surrounding a school. We split the sample into "high" and "low" income schools, where high corresponds to schools that are in counties with median household incomes above $30,000 in the 1990 census, and low corresponds to schools in counties below this level. Again, controlling for income levels does not change the basic patterns observed in the preference and meeting biases, but we do see significant differences in preference biases between high and low income schools, although the differences in meeting biases do not differ significantly. We see less preference bias for some races (Hispanics and Whites) in higher income schools and the reverse for Blacks.

[‡]Larger schools also offer some advantages, as there may be economies of scale and it may also be easier to draw more racially balanced populations in larger districts. The significant increases in meeting biases in large schools, however, suggest that either one might want to create schools within schools, or understand the factors leading to increased meeting biases and homophily in larger schools.

Table 1. Estimation of preference and meeting biases

Preference parameter	α	γ_{Asians}	γ_{Blacks}	$\gamma_{Hispanics}$	γ_{Whites}	γ_{Others}
Estimated value	0.55	0.90	0.55	0.65	0.75	0.90
Meeting parameter		β_{Asians}	β_{Blacks}	$\beta_{Hispanics}$	β_{Whites}	β_{Others}
Estimated value		7	7.5	2.5	1	1

Preference bias on a grid of step 0.05, from 0.20 to 0.85 for α and from 0.40 to 1 for each of the γ; meeting bias on a grid of step 0.5 from 1 to 9.

Materials and Methods

Here we describe the model and empirical analysis, referring to *SI Text* for technical details.

In our model a population of agents is partitioned into K different groups, where group definitions delineate the relevant characteristics that agents care about in forming friendships. Let w_i denote the fraction of the population that is of type i. Agents have preferences over the number of friends that they have from their own group, henceforth referred to as "same-type" friends, and from other groups, henceforth referred to as "different-type" friends. Let s_i and d_i denote the number of same-type and different-type friends that a representative agent in group i has, and let $t_i = s_i + d_i$ be the agent's total number of friends. The agent's preferences are represented by a utility function,

$$U_i(s_i, d_i) = (s_i + \gamma_i d_i)^{\alpha}, \qquad [1]$$

where both γ_i and α lie between 0 and 1, so that U_i is increasing in both s_i and d_i. The function U_i measures the utility or satisfaction drawn from one's friendships. The parameter γ_i captures the bias in preferences, with $\gamma_i < 1$ indicating that different-type friends are valued less than same-type friends. The parameter α captures diminishing returns to friendships overall: doubling the number of each type of friends that an agent has results in less than double the utility. Finally, agents only distinguish between same-type and different-type friends, as is roughly consistent with empirical evidence (20–23).

Friendship formation takes place via a meeting process in which agents randomly meet potential friends, perhaps in a biased manner. The meeting process can be thought of as a sort of "party" where agents come to the party and randomly meet other agents and then eventually leave the party once the benefit from meeting more friends no longer exceeds the opportunity cost ($c > 0$) of time and resources of staying at the party. The relevant meeting parameter from an agent's decision perspective is the expected rate at which he/she will meet same-type versus different-type friends at the party. For type i, these are denoted by q_i and $1 - q_i$, respectively. If preferences are biased so that $\gamma_i < 1$ and same-type friendships yield higher marginal values than different-type friendships, then a higher matching rate q_i provides higher incentives to form additional friendships. In this case, groups that meet their own types at higher rates (face higher q_i) at the party will choose to stay at the party longer and thus form more friendships per capita than groups that meet their own types at lower rates. The meeting rates for different races and total numbers of friendships are both observable in the data, allowing us to identify the preference parameters. Note also that choice and chance feed back upon each other: given biased preferences groups accounting for a larger share of the population will choose to stay at the party longer (i.e., socialize more) and so end up forming an even greater portion of people at the party, thus making it even more attractive for their types and less attractive for other types. This feedback does not prevent us from identifying preference and meeting biases, as described below.

Solving the model requires determining the meeting probabilities. If meetings follow a uniform random process (in which case we talk of an "unbiased" meeting process), agents meet each other in proportion to their relative stocks (i.e., proportions at the party), so that $q_i = \frac{M_i}{\sum_k M_k}$, where $M_i = w_i t_i$ is the stock of agents of type i. Biases in meetings, such that agents meet same friends at higher rates than their relative stocks, are captured by

$$q_i = \left(\frac{M_i}{\sum_k M_k}\right)^{1/\beta_i}, \qquad [2]$$

where $\beta_i > 1$ is the bias that each type has toward itself in the meeting process, and $\beta_i = 1$ is the case of unbiased process. For instance, if a group comprises half of the meeting pool and has a bias of $\beta_i = 2$, then it would meet itself at a rate of $(0.5)^{1/2}$ or ∼0.7, while if $\beta_i = 3$ this rises to ∼0.8, and at $\beta_i = 6$ is ∼0.9. Given that the stocks of different types in the meeting process must sum to one, $\sum_i \frac{M_i}{\sum_k M_k} = 1$, it follows that

Table 2. *F*-statistics of constrained calibrations, compared with the unconstrained calibrations in which every race has a different parameter

	Preference bias, γ			Meeting bias, β		
	F-statistic	95%	99%	*F*-statistic	95%	99%
Asian = Black	9.93**	3.96	6.97	0.04	3.96	6.97
Asian = Hispanic	8.17**	"	"	4.95*	"	"
Asian = White	2.65	"	"	47.97**	"	"
Black = Hispanic	1.56	"	"	19.31**	"	"
Black = White	10.43**	"	"	124.5**	"	"
Hispanic = White	3.43	"	"	23.34**	"	"
All = 1	42.61**	2.33	3.26	220.6**	2.33	3.26
All =	6.10**	2.49	3.57	51.25**	2.49	3.57

*Significance above a 95 percent level; **significance above a 99 percent level.

$$\sum_i q_i^{\beta_i} = 1. \qquad [3]$$

We also impose conditions that relate meeting rates across races, since if a person of type i is meeting a person of type j then the converse is also true, as described in *SI Text*. The meeting process is illustrated in Fig. 2.

Using this model, we estimate the preference bias and meeting bias parameters from the data. If an agent forms t_i friendships when a fraction of q_i are of same-type, then the resulting utility (including costs of time in the meeting process) is $(t_i q_i + \gamma_i(1 - q_i)t_i)^{\alpha} - ct_i$. Thus, utility optimization (see *SI Text* for details) implies that for every type i the following first order necessary condition holds:

$$t_i = \left(\frac{\alpha}{c}\right)^{\frac{1}{1-\alpha}} (\gamma_i + (1-\gamma_i)q_i)^{\frac{\alpha}{1-\alpha}}. \qquad [4]$$

From Eq. **4**, we see that γ_i dictates how sensitive the total number of friends of a given agent is to changes in the odds of meeting a same-type agent. If there is not much preference bias (i.e., a γ_i near 1), then the number of friendships formed is relatively insensitive to the rate at which same-type friends are met. In contrast, if preferences are heavily biased toward own-type (i.e., γ_i is lower), then the number of friendships formed will be very sensitive to the rate at which same-type friends are met. This is a key to the identification: the sensitivity of the number of friendships formed by a given type of agent to the same-type meeting rate identifies the bias in preferences. The identification of the meeting biases can then be found from Eq. **3**. To see why it identifies the meeting biases, note that it implicitly keeps track of the extent to which the relative fraction of types met (the q_is) differ from the relative stocks at the 'party' (the w_is weighted by the t_is).

As described in *SI Text*, we allow for individual idiosyncrasies in preferences and other perturbations, by allowing Eq. **4** to only hold up to an individual error term, and we also allow costs to vary by school. Applying condition **4** to any pair of types i and j, we can eliminate the cost term (which is unobserved in the data) and obtain the following equation:

$$t_i(\gamma_j + (1-\gamma_j)q_j)^{\frac{\alpha}{1-\alpha}} = t_j(\gamma_i + (1-\gamma_i)q_i)^{\frac{\alpha}{1-\alpha}}. \qquad [5]$$

Note that t_i, t_j, and q_i, q_j are available data, since t_i is the number of friendships on average by type i students, and $q_i = s_i/(s_i + d_i)$ is the average percent of friendships formed by type i students that are of the same-race. We thus estimate α and the γ_i's by minimizing the errors in Eq. **5**. Then, each choice of values for α, γ_{Asians}, γ_{Blacks}, $\gamma_{Hispanics}$, γ_{Whites}, and γ_{Others} leads to a difference between the right hand side and the left hand side of Eq. **5**. Weighting schools to correct for their size and hence the variance in errors which are coming from individual choices (as described in detail in *SI Text*) yields a sum of squared errors for each choice of α and γ's. The statistical analysis reported below is based on a presumption that the errors follow a Normal distribution, and in *SI Text*, Fig. S2, we verify that the realized distribution of errors does not differ significantly from a Normal distribution. We search over a grid to find parameters that minimize this weighted sum of squared errors. To estimate the β parameters, we follow the same technique based on Eq. **3**. Results are reported in Table 1.

The patterns across race differ significantly. Asian students are the least biased in their preferences over racial mixes, having the highest γ at 0.9, but

ECONOMIC SCIENCES

Table 3. Preference and meeting biases when allowed to vary by school size

	α	γ_a	γ_b	γ_h	γ_w	γ_o	RSS	*F*	95% thresh.	99% thresh.
Ignoring size	0.55	0.90	0.55	0.65	0.75	0.90	4704	—	—	—
Small schools	0.65	0.90	0.75	0.80	0.80	0.90	1685	—	—	—
Large schools	0.55	0.85	0.40	0.45	0.65	0.85	2531	—	—	—
	Total error small + large						4216	1.39	2.23	3.06
	β_a	β_b	β_h	β_w	β_o	—	RSS	*F*	95% thresh.	99% thresh.
Ignoring size	7.0	7.5	2.5	1.0	1.0	—	1.7265	—	—	—
Small schools	6.5	6.0	2.0	1.0	1.0	—	0.9406	—	—	—
Large schools	3.0	9.0	6.5	1.0	1.0	—	0.3688	—	—	—
	Total error small + large					—	1.3094	4.71**	2.34	3.28

Table 4. Preference and meeting biases when allowed to vary by the school's county median household income level (low is <30,000 dollars in 1990 census and high is above)—for 78 of the 84 schools for which we have income data

	α	γ_a	γ_b	γ_h	γ_w	γ_o	RSS	*F*	95% thresh.	99% thresh.
Ignoring income	0.55	0.95	0.55	0.65	0.75	0.90	4255	—	—	—
Low income schools	0.60	1.0	0.50	0.50	0.65	0.95	1541	—	—	—
High income schools	0.35	1.0	0.40	0.70	0.90	0.75	1703	—	—	—
	Total error low + high						3244	3.43**	2.23	3.06
	β_a	β_b	β_h	β_w	β_o	—	*RSS*	*F*	95% thresh.	99% thresh.
Ignoring income	7.5	7.5	2.5	1.0	1.0	—	1.652	—	—	—
Low income schools	2.0	8.0	3.0	1.0	1.0	—	0.8699	—	—	—
High income schools	5.0	4.5	4.0	1.0	1.0	—	0.7485	—	—	—
	Total error low + high					—	1.618	0.28	2.34	3.28

they face a very substantial meeting bias with a β parameter of 7. Blacks exhibit the greatest bias in both preferences and meetings, with a γ of 0.55 and a β of 7.5. Hispanics are intermediate in terms of both biases, while Whites have an intermediate preference bias and no meeting bias.

To check that these differences reflect a systematic heterogeneity and not the mere effect of randomness, we run a series of F-tests on the above calibrations, as described in *SI Text* and Tables S4–S7. A summary of the results appears in Table 2. The hypothesis that all of the races have unbiased preferences is rejected well above the 99% confidence level. Moreover, Asian and White preference biases do not differ significantly from each other, and the same holds for Blacks and Hispanics, but Blacks and Hispanics both differ significantly from both Asians and Whites. Similar patterns are found in the meeting biases, but this time with Asians and Blacks having indistinguishable meeting biases but both differing significantly from Hispanics and Whites.[§]

Beyond the analysis above, we also control for the effect of school size on the estimated parameters, as well as income. We analyze the effect of school size by re-estimating the model, when the schools are divided into a category of large schools (with >1000 students) and small schools (with <1000 students). We can then compare the estimated preference and meeting biases by school size. As reported in Table 3, larger schools exhibit significantly higher meeting biases (at the 99% level) than small schools, but differences in preference biases between small and large schools are insignificant (even at the 90% confidence level). In particular, meeting biases are greater for Blacks and Hispanics in large compared to small schools, but lesser for Asians in large compared to small schools. We also checked whether the patterns in biases are sensitive to the median household income level in the school's county, since it turns out that income shows very little correlation with school size in these data. Again, we divided schools into two categories, those in which the county level median household income level was above \$30,000 in the 1990 census, and those for which it was below that level. As reported in Table 4, in this case preference biases differ significantly (at the 99% level) across the high and low income-level schools, but the meeting biases do not differ significantly (even at the 90% confidence level) across the high and low income-level schools. Here we see Hispanic and White exhibiting more bias in preferences (i.e., having lower γ's) in low income schools and there is more bias overall in low income schools (when estimating a single bias parameter), although Blacks exhibit slightly more preference bias in high income schools.

Further study is needed to understand the sources of preference and meeting biases, why they differ across races, and why they are correlated with school size and median income. Moreover, the racial categorization here is quite coarse,[¶] and many other attributes can also affect friendship formation. In addition, we note that our analysis might even underestimate preference biases, since meeting biases could incorporate some aspects of choice, and since meetings are partly endogenous given that students have some choice as to which clubs to join, which sports to participate in, which parties to go to, and so forth. Indeed, there is a positive correlation between inbreeding homophily and the number of clubs and athletic activities that a school has (a correlation of 0.26 which is significant at the 99% confidence level).[‖]

Beyond this, it would be interesting to extend the revealed preference sort of techniques used here to incorporate the sort of network induced homophily effects found by Kossinets and Watts (24, 25), who show the importance of proximity in an existing social network for the formation of new friendships.**

Finally, it is worth emphasizing that the "revealed preference" techniques applied here can be employed in many other settings; by specifying an

[§]In *SI Text* we also report estimation of these biases when restricting attention to groups with minimal weights in the population, as well as estimations of the meeting biases by estimating Eq. **2** on a race-by-race basis, rather than as a joint estimation procedure as we use under Eq. **3**. These provide biases that are similarly significantly above 1, but with some compression and variation in the differences across races.

[¶]The coarseness could lead the preference biases of some racial groups to be underestimated. For example, the "Asian" group includes Chinese, Japanese, Korean, Vietnamese, Indian, and many other populations under one umbrella. To the extent that their racial preferences differ by the finer ethnic categorizations, it could be that this leads their friendship patterns with other categories to look similar to the same group, and so leads to a higher γ parameter than would be estimated if data were available on finer categorizations. This might account for the significant differences between Asian preference bias and some of the other groups' biases.

[‖]We consider activities that have at least two student members and discard the few students who claimed to be involved in 20 or more clubs or sports with some having claimed to be involved in all possible activities.

**It is worth emphasizing that the terms "choice homophily" and "preferences," as used in the empirical literature on homophily, generally refer to some conditional likelihood, or log odds ratios, of forming ties based on some characteristics, rather than explicit modeling of a maximization of a utility function as we have done here. Thus, a hybrid of the sort of model analyzed here together with the network evolution and constraints that network structures place on the meeting process as analyzed in refs. 24 and 25 could be quite valuable.

appropriate preference structure and a process through which agents might meet each other, and then identifying various parameters through the co-movements of the fraction of agents of various types and their patterns of relationships.

ACKNOWLEDGMENTS. We thank James Moody for making available the network data from the Add Health dataset. We gratefully acknowledge financial support from the National Science Foundation (Grant SES-0647867) and the Italian Ministry of Research (Grant 2007TKLTSR). This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. We thank Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516–2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

1. Moody J (2001) Race, school integration, and friendship segregation in America. *Am J Sociol* 107:679–716.
2. Patacchini E, Zenou Y (2006) Racial identity and education. *IZA Discuss Pap*, Paper 2046.
3. Currarini S, Jackson MO, Pin P (2009) An economic model of friendship: homophily, minorities and segregation. *Econometrica* 77:1003–1045.
4. Fowler JH, Dawes CT, Christakis NA (2009) Model of genetic variation in human social networks. *Proc Natl Acad Sci USA* 106:1720–1724.
5. Lazarsfeld PF, Merton RK (1954) Friendship as a social process: a substantive and methodological analysis, in *Freedom and Control in Modern Society*, ed Berger M (Van Nostrand, New York), pp 18–66.
6. McPherson JM, Smith-Lovin L (1987) Homophily in voluntary organizations: status distance and the composition of face-to-face groups. *Am Sociol Rev* 52:370–379.
7. Jackson MO (2008) *Social and Economic Networks* (Princeton Univ Press, Princeton).
8. Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities, in *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York), pp 160–168.
9. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence based contagion from homophily driven diffusion in dynamic networks. *Proc Natl Acad Sci USA*, 10.1073/pnas.0908800106.
10. Schuman H, Steeh C, Bobo LD, Krysan M (1997) *Racial Attitudes in America: Trends and Interpretations* (Harvard Univ Press, Cambridge, MA).
11. Mas-Colell A, Whinston M, Green J (1995) *Microeconomic Analysis* (Oxford Univ Press, Oxford).
12. Coleman J (1958) Relational analysis: the study of social organizations with survey methods. *Hum Organ* 17:28–36.
13. Berry CR, West MR (2008) Growing pains: the school consolidation movement and student outcomes. *J Law Econ Org*.
14. Coladarci T, Cobb CD (1996) Extracurricular participation, school size, and achievement and self-esteem among high school students: a national look. *J Res Rural Educ* 12:92–103.
15. Khmelkov VT, Hallinan MT (1999) Organizational effects on race relations in schools. *J Soc Issues* 55:627–645.
16. Morgan DL, Alwin DF (1980) When less is more: school size and student social participation. *Soc Psychol Q* 43:241–252.
17. Ornstein AC (1990) School size and effectiveness: policy implications. *Urban Rev* 22: 239–245.
18. Slate JR, Jones CH (2005) Effects of school size: a review of the literature with recommendations. *Essays Educ* 13.
19. Stearns E (2004) Interracial friendliness and the social organization of schools. *Youth Soc* 35:395–419.
20. Marsden PV (1987) Core discussion networks of Americans. *Am Sociol Rev* 52:122–131.
21. Marsden PV (1988) Homogeneity in confiding relations. *Soc Networks* 10:57–76.
22. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27:415–444.
23. Goodreau SM, Kitts JA, Morris M (2009) Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46:103–125.
24. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311:88–90.
25. Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network. *Am J Sociol* 115:405–450.

ECONOMIC SCIENCES