NBER WORKING PAPER SERIES

SOCIAL CAPITAL II:
DETERMINANTS OF ECONOMIC CONNECTEDNESS

Raj Chetty
Matthew O. Jackson
Theresa Kuchler
Johannes Stroebel
Nathaniel Hendren
Robert B. Fluegge
Sara Gong
Federico Gonzalez
Armelle Grondin
Matthew Jacob
Drew Johnston
Martin Koenen
Eduardo Laguna-Muggenburg
Florian Mudekereza
Tom Rutter
Nicolaj Thor
Wilbur Townsend
Ruby Zhang
Mike Bailey
Pablo Barberá
Monica Bhole
Nils Wernerfelt

Working Paper 30314
http://www.nber.org/papers/w30314

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2022

R. Chetty, M. Jackson, T. Kuchler, and J. Stroebel contributed equally to this work. They were the joint principal investigators on this project and designed the study, supervised all analyses, analyzed data, and wrote the paper. R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, and R. Zhang

analyzed data, prepared figures, and provided conceptual contributions to the study. M. Bailey led the collaboration between the external researchers and the Meta Research Team. N. Hendren, E. Laguna-Muggenburg, P. Barbera, M. Bhole, and N. Wernerfelt provided intellectual input and edited drafts of the manuscript. We are grateful to J. Friedman, M. Gentzkow, E. Glaeser, R. Putnam, B. Sacerdote, A. Shleifer, and numerous seminar participants for helpful comments, G. Crowne, T. Harris, A. Kim, J. Sun, V. Weiss-Jung, and A. Zheng for excellent research assistance, A. Hiller and S. Oppenheimer for project management and content development, S. Halvorson, R. Korzan, C. Shram, and M. Wong of Darkhorse Analytics for creating the data visualization platform, S. Vadhan for help in developing the differential privacy methods used in this paper, and the Meta Research Team for their support. This research was facilitated through a research consulting agreement between Chetty, Jackson, Kuchler, and Stroebel and Meta Platforms, Inc. Jackson is an external faculty member of the Santa Fe Institute. The work was funded by the Bill & Melinda Gates Foundation, the Overdeck Family Foundation, Harvard University, and the National Science Foundation (under grants SES-1629446 and SES-2018554 issued to Jackson in his academic capacity at Stanford). Opportunity Insights also receives core funding from other sponsors, including the Chan Zuckerberg Initiative, the Robert Wood Johnson Foundation, and the Yagan Family Foundation. CZI is a separate entity from Meta, and CZI funding to Opportunity Insights was not used for this research. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders. In 2018, T.K. and J.S. received an unrestricted gift from Facebook to NYU Stern. T. Kuchler, J. Stroebel, S. Gong, and F. Mudekereza. are contract affiliates through Meta's contract with PRO Unlimited. F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, T. Rutter, N. Thor, W. Townsend, and R. Zhang are contract affiliates through Meta's contract with Harvard University. Meta Platforms, Inc. did not dispute or influence any findings or conclusions during their collaboration on this research. This work was produced under an agreement between Meta and Harvard University specifying that Harvard shall own all intellectual property rights, titles and interests (subject to the restrictions of any journal or publisher of the resulting publication(s)).

Social Capital II: Determinants of Economic Connectedness
Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt
NBER Working Paper No. 30314
July 2022
JEL No. R0

ABSTRACT

Low levels of social interaction across class lines have generated widespread concern and are associated with worse outcomes, such as lower rates of upward income mobility. Here, we analyze the determinants of cross-class interaction using data from Facebook, building upon the analysis in the first paper in this series. We show that about half of the social disconnection across socioeconomic lines—measured as the difference in the share of high-socioeconomic status (SES) friends between low- and high-SES people—is explained by differences in exposure to high- SES people in groups such as schools and religious organizations. The other half is explained by friending bias—the tendency for low-SES people to befriend high-SES people at lower rates even conditional on exposure. Friending bias is shaped by the structure of the groups in which people interact. For example, friending bias is higher in larger and more diverse groups and lower in religious organizations than in schools and workplaces. Distinguishing exposure from friending bias is helpful for identifying interventions to increase cross-SES friendships (economic connectedness). Using fluctuations in the share of high-SES students across high school cohorts, we show that increases in high-SES exposure lead low-SES people to form more friendships with high-SES people in schools that exhibit low levels of friending bias. Hence, socioeconomic integration can increase economic connectedness in communities where friending bias is low. In contrast, when friending bias is high, increasing cross-SES interaction among existing members may be necessary to increase economic connectedness. To support such efforts, we release privacy-protected statistics on economic connectedness, exposure, and friending bias for each ZIP code, high school, and college in the U.S. at www.socialcapital.org.

Raj Chetty
Department of Economics
Harvard University
Littauer 321
Cambridge, MA 02138
and NBER
chetty@fas.harvard.edu

Theresa Kuchler
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012 and
NBER
tkuchler@stern.nyu.edu

Matthew O. Jackson
Department of Economics
Stanford University
Stanford, CA 94305-6072
and External Faculty of the Santa Fe Institute
jacksonm@stanford.edu

Johannes Stroebel
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012 and
NBER
johannes.stroebel@nyu.edu

Nathaniel Hendren
Harvard University
Department of Economics
Littauer Center Room 235
Cambridge, MA 02138
and NBER
nhendren@gmail.com

Robert B. Fluegge
Harvard University
rfluegge@g.harvard.edu

Sara Gong
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
sgong@stern.nyu.edu

Federico Gonzalez
Harvard University
federico_gonzalez@g.harvard.edu

Armelle Grondin
Opportunity Insights
1280 Massachusetts Ave
Second floor
Cambridge, MA 02138
agrondin@fas.harvard.edu

Matthew Jacob
Harvard University
mjacob@g.harvard.edu

Drew Johnston
Harvard University
drewjohnston@g.harvard.edu

Martin Koenen
Harvard University
martin_koenen@fas.harvard.edu

Eduardo Laguna-Muggenburg
Grammarly
475 Sansome St.
San Francisco, CA 94111
edu.laguna@stanford.edu

Florian Mudekereza
Opportunity Insights
1280 Massachusetts Ave
Second floor
Cambridge, MA 02138
florianmudekereza@g.harvard.edu

Tom Rutter
Opportunity Insights
1280 Massachusetts Ave
Second floor
Cambridge, MA 02138
trutter@g.harvard.edu

Nicolaj Thor
Opportunity Insights
1280 Massachusetts Ave
Second floor
Cambridge, MA 02138
nthor@g.harvard.edu

Wilbur Townsend
Harvard University
wilbur.townsend@gmail.com

Ruby Zhang
Harvard University
rzhang15@g.harvard.edu

Mike Bailey
Meta Platforms
1 Hacker Way
Menlo Park, CA 94025
mcbailey@fb.com

Pablo Barberá
Meta Platforms
1 Hacker Way
Menlo Park, CA 94025
pbarbera@fb.com

Monica Bhole
Meta Platforms
1 Hacker Way
Menlo Park, CA 94025
mbhole@fb.com

Nils Wernerfelt
Meta Platforms
1 Hacker Way
Menlo Park, CA 94025
nilsw@fb.com

Website for Data Visualization and Download is available at www.socialcapital.org
Replication Code is available at www.opportunityinsights.org/data

I Introduction

Many researchers and policy makers have raised concerns that societies around the world are fragmented and polarized, with little interaction across racial, political, and class lines (Doob 2019; Fischer and Mattson 2009; Gentzkow and Shapiro 2011; Putnam 2016; Smith et al. 2014). In addition to being of potential concern in its own right, a lack of interaction between different types of people is associated with worse economic and social outcomes (Alesina et al. 1999; Putnam 2016, 2020). For example, in the preceding paper in this series (Chetty et al. 2022), we used data on social networks from Facebook to show that communities where low-socioeconomic status (SES) people interact less with high-SES people exhibit less upward income mobility across generations.

In light of these findings, in this paper, we analyze the determinants of social interactions across class lines (economic connectedness). Why do higher SES people tend to have more high-SES friends than low-SES people do? Building on prior work (e.g., Feld 1982, Zeng and Xie 2008, Currarini et al. 2009), we distinguish two channels that can generate differences in an individual's share of high-SES friends: (i) differences in high-SES *exposure,* the share of high-SES members in the groups to which low- vs. high-SES people belong (e.g., their schools or religious organizations), and (ii) differences in *friending bias*, the rate at which people befriend high-SES individuals conditional on the share of high-SES members in the group (i.e., conditional on exposure).[1] Distinguishing these two channels is critical for developing interventions to increase economic connectedness: if differences in exposure are central, then efforts to increase socioeconomic integration in neighborhoods and schools may be the key to increasing connectedness; in contrast, if friending bias is central, one must instead focus on how to increase social interaction across class lines *within* neighborhoods and schools.

We quantify the relative importance of exposure and friending bias in explaining differences in economic connectedness using data on the social networks of 70.3 million U.S. Facebook users between the ages of 25–44.[2] We measure each individual's socioeconomic status by combining several proxies, such as neighborhood incomes and educational attainment, as described in Chetty et al. (2022). We allocate each individual's friendships to the groups in which they were formed—such as high schools, colleges, religious groups, workplaces, and neighborhoods—using information from Facebook profiles

[1] We use the term friending "bias" in a statistical sense, to denote biased sampling from the pool of available peers. Friending bias should not be interpreted as a normative statement about individuals' preferences; indeed, as we show below, it could result from structural or institutional factors, such as tracking within schools.

[2] Facebook data have been used to study the effects of social networks on a variety of outcomes in prior work, ranging from home purchasing and mortgage decisions (Bailey et al. 2018b, 2019), to labor market outcomes (Gee 2018; Gee et al. 2017), to trade flows (Bailey et al. 2018a, 2020b), to EITC claiming behavior (Wilson 2020); see Chetty et al. (2022) for additional examples. As in this prior work, we use social network data to proxy for "real world" friendships rather online interactions per se; as a result, our analysis does not shed light on the effects of online social networks themselves on exposure and friending bias.

and group memberships.[3] Combining these data, we construct privacy-protected measures of exposure and friending bias for every group—i.e., for each high school, college, ZIP code, and county—in the United States.

We find that about half of the difference in the fraction of high-SES friends between low- and high-SES individuals is explained by differences in exposure. For example, high-SES individuals attend high schools that are disproportionately attended by other high-SES people. The other half of the difference is largely explained by friending bias: within each group, high-SES individuals are more likely than low-SES individuals to form friendships with high-SES peers. Hence, exposure is a necessary but not sufficient condition for cross-SES interaction: even when the rich and poor live in the same neighborhoods and attend the same schools, low-SES people still have substantially fewer high-SES friends than high-SES people do. Differences in exposure and friending bias are equally predictive of differences in upward income mobility across areas, suggesting that it is cross-SES *interaction* rather than simply physical integration that predicts economic mobility.

Having established that both exposure and friending bias contribute to differences in economic connectedness and its relationship with economic mobility, we turn to analyzing their determinants. The determinants of exposure have been widely analyzed in the literature on income segregation across neighborhoods, schools, and other groups (Chetty et al. 2020; Davis et al. 2019; Owens et al. 2016; Reardon and Bischoff 2011; Reardon and Owens 2014). We therefore focus here on the determinants of friending bias, exploiting our social network data's unique ability to shed light on the degree to which different types of people actually *interact* within groups.[4]

Similar to exposure—which depends on institutional and policy choices such as college admissions policies and zoning laws—friending bias varies systematically across settings and appears to be influenced by the structure of institutions. For example, friending bias is much lower in religious institutions than in other settings: the friendships low-SES people make in their religious groups are more likely to cut across class conditional on exposure than the friendships they make in their schools or neighborhoods.[5] Low-SES people also exhibit more friending bias in larger and more diverse groups. In sum, both components of economic connectedness—exposure and friending bias—are determined by structural features of societies that shape social interaction, potentially creating scope for policy changes to

[3]The settings in which friendships are made differ by SES: lower-SES people make a larger share of their friendships in the neighborhoods where they live, while higher-SES individuals make more of their friendships in college, perhaps explaining why the economic outcomes of the poor vary more sharply across neighborhoods than those of the rich (Chetty et al. 2018).

[4]The capacity to identify interaction rather than exposure (geographic proximity) is a key distinction between the present study and recent work that measures experienced segregation using location data from mobile devices (e.g., Athey et al. 2020; Dong et al. 2020; Levy et al. 2020; Reme et al. 2022; Wang et al. 2018).

[5]Despite having little friending bias, religious groups currently do not have high levels of economic connectedness because they tend to be highly segregated by income. However, efforts to increase economic integration in religious groups could be particularly effective in increasing cross-SES connections if bias remains low as integration increases.

foster greater cross-class interactions.

In the last part of the paper, we demonstrate how our institution-specific estimates of exposure and friending bias can inform such policy interventions. Focusing on connectedness in high schools, we estimate the causal effects of socioeconomic integration on economic connectedness by exploiting fluctuations in the share of high-SES students across cohorts coupled with a regression discontinuity design based on birth date cutoffs for school entry. In schools with low friending bias in other cohorts, quasi-random assignment to a cohort with more high-SES students produces many more friendships between low- and high-SES students. However, in schools with high friending bias, the same change in exposure to high-SES peers leads to significantly fewer additional friendships between low- and high-SES students.

These quasi-experimental results suggest that economic integration can increase economic connectedness in groups that currently have low levels of friending bias. However, in groups with higher friending bias, it may be more fruitful to address the determinants of this bias, for example through programs or curricular changes that foster greater cross-SES interaction within schools or the creation of new settings for people to interact (Alan et al. 2020; Pallais and Anders 2021; Rohrer et al. 2021). Such efforts may be particularly important in environments that have the type of large, diverse populations that we find are otherwise prone toward segregation and friending bias. To support efforts to increase economic connectedness and further study its determinants, we have made privacy-protected statistics on exposure and friending bias by ZIP code, school, and college freely accessible for download and visualization at www.socialcapital.org.[6]

The paper is organized as follows. In Section II, we define the parameters we seek to identify in our empirical analysis by describing how we decompose economic connectedness into exposure and friending bias. Section III presents our findings on the relative importance of exposure and friending bias in explaining economic connectedness and economic mobility. In Section IV, we present the estimates of exposure and friending bias by high school and college, analyze their association with institutional characteristics, and demonstrate how these estimates can be helpful in targeting interventions by analyzing the causal effects of integration in high schools. Section V discusses the implications of the findings. Section VI presents details on our data, sample definitions, and our methodology. Additional technical details on data, methods, and supplementary analyses are available in the Supplementary Information.

[6]To protect user privacy, we use methods from the differential privacy literature to add noise to these statistics while maintaining a high level of statistical reliability (Dwork et al. 2006, Chetty and Friedman 2019).

II Determinants of Economic Connectedness

Following Chetty et al. (2022), we measure the degree of cross-class interaction—which we term *economic connectedness*—as the share of above-median-SES ("high-SES") friends among below-median-SES ("low-SES") people divided by 50%, in order to quantify the average degree of under-representation of high-SES friends among low-SES people. A value of 0 for economic connectedness (EC) implies that a network has no connections between low- and high-SES people, while a value of 1 implies that low-SES people have an equal number of low- and high-SES friends. On average, EC is 0.776 for low-SES people in the U.S. (Chetty et al. 2022), i.e., high SES friends are *under-represented* by 22.4% ($1 - 0.776 = 0.224$) among low-SES individuals relative to the high-SES share in the population. Our goal is to determine the factors that generate this 22.4% under-representation of high-SES friends.

Within any group where friendships are made—such as a specific high school or neighborhood—the rate at which low-SES people become friends with high-SES people depends on two factors: (i) high-SES *exposure* (Allport et al. 1954; Blau and Schwartz 1984), the share of high-SES members in the group and (ii) *friending bias* (Currarini et al. 2009), the rate at which people befriend high-SES individuals conditional on the share of high-SES members in the group (i.e., conditional on exposure). For example, in the context of schools, low-SES students may have fewer high-SES friends because they attend schools with few high-SES students (i.e., schools with low exposure) or because they are less likely to befriend high-SES students even within their schools (that is, their schools have high friending bias).

We measure exposure in a group as the share of above-median-SES individuals in the group multiplied by two, so that exposure is equal to 1 for a group with 50% above-median-SES individuals (see Section VI.D for algebraic definitions of this and subsequent terms in this section). Exposure is below 1 for groups that have a below-average share of high-SES individuals and above 1 for groups that have an above-average share of high-SES individuals.

We define a person's friending bias in each group as one minus the share of friends they make in that group who have high SES divided by the share of *people* in the group with high SES. If friendships were formed at random—and if high- and low-SES people made the same number of friends—then individuals' share of high-SES friends in a given group would equal the share of high-SES individuals who belong to the group and friending bias would be equal to 0. Friending bias greater than 0 implies a lower probability of making high-SES friends than if friendships were formed at random within a given group.

In practice, high-SES individuals make more friends than low-SES individuals do on average (Chetty et al. 2022). Maintaining this difference in the number of total friends, in a setting with no homophily

by SES (i.e., where low- and high-SES individuals have the same probability of befriending a given high-SES person), friending bias would be negative; that is, low-SES individuals would have more high-SES friends than the share of high-SES people in the population. Quantitatively, given the empirically observed difference in the number of friends by SES, we would expect friending bias to be -11% in the absence of homophily (see Section VI.D).

Note that the distinction between exposure and friending bias depends on the level at which one measures exposure, and is therefore a policy-dependent rather than conceptual distinction. For example, friending bias in schools may arise from differences in high-SES exposure within schools due to tracking of students into different classrooms. Nevertheless, measuring exposure and friending bias at the school level has policy relevance because many interventions to increase socioeconomic integration, such as busing or changes in school attendance boundaries, focus on integration across rather than within schools (Supplementary Information C.1). Relatedly, the term friending "bias" should be interpreted in a statistical sense—denoting biased sampling from the pool of available peers—rather than as a normative statement about individuals' preferences, as bias may be the result of institutional factors such as within-school tracking.

In addition to levels of exposure and friending bias *within* groups, economic connectedness also depends upon *where* people make friends. For example, low-SES people are less likely to attend college and hence make fewer friends in college than in other settings. Since colleges tend to have many high-SES people and high levels of economic connectedness, low college attendance rates thus reduce low-SES people's economic connectedness (holding fixed exposure and friending bias at all colleges).

We measure exposure, friending bias, and the share of friends individuals make in six settings, which comprise the most common places where people make friends (Jeffres et al. 2009; Oldenburg 1999): high schools, colleges, religious groups, recreational groups, workplaces, and neighborhoods. We estimate these measures separately by group (e.g., separately for each high school in the U.S.) in each of these six settings using privacy protected data from Facebook (see Section VI.J). As in Chetty et al. (2022), we focus on Facebook users aged between 25 and 44 who reside in the United States, were active on the Facebook platform at least once in the prior 30 days, have at least 100 U.S.-based Facebook friends, and have a non-missing ZIP code. Here, we further restrict attention to individuals for whom we can allocate at least one friendship to the group in which it was formed, using the approach described in Section VI.B. The resulting sample consists of 70.3 million Facebook users, corresponding to 82% of the U.S. population between ages 25–44. Based on comparisons to nationally representative surveys and other supplementary analyses, we find that our Facebook analysis sample is reasonably representative of the national population (see Section VI.C and Chetty et al. 2022).

We use the Facebook data to obtain information on friendships, locations (ZIP code and county), own and parental socioeconomic status, and group memberships; we describe these variables in detail in Section VI.B. To capture the varied definitions of socioeconomic status used in prior work by Karl R. White (1982b), we compute socioeconomic status by combining several measures of SES, such as average incomes in an individual's neighborhood and levels of educational attainment. We combine these measures of SES into a single index using a machine learning algorithm described in Section VI.B and discussed further and validated against external measures in Chetty et al. (2022). We allocate each individual's friendships to the groups in which they were formed using information from Facebook profiles and group memberships (see Section VI.B).

III Determinants of Economic Connectedness: Exposure vs. Friending Bias

III.A Exposure and Friending Bias by Setting

We begin by analyzing how rates of friendship formation, economic connectedness, exposure, and friending bias vary across settings.

Figure 1 shows how the share of friends that an individual makes in each setting varies with their SES rank. For each SES ventile, it plots the average proportion of friends made in each setting, divided by the overall proportion of friends made in that setting across all SES ventiles. The lowest-SES individuals make about four times as large a share of their friends in their neighborhoods (residential ZIP codes) as the highest-SES individuals do. Conversely, high-SES individuals make a far larger share of their friends in college than low-SES individuals do, primarily because high-SES individuals are much more likely to attend college. Neighborhoods thus play a larger role in defining low-SES individuals' social communities, perhaps explaining why where one lives matters more for the economic and health outcomes of the poor than the rich (Chetty et al. 2018, 2016).

Figure 2a shows how economic connectedness (EC) varies across the six settings for below-median-SES and above-median-SES people. For each SES category, we define the setting-specific EC as two times the average share of high-SES friends (among friends made in that setting). EC for low-SES people is highest in colleges and lowest in their residential neighborhoods. However, even in colleges, low-SES people are much less likely to befriend high-SES people than high-SES people are. To understand why, we next examine rates of high-SES exposure and friending bias in each setting.

Figure 2b plots average exposure to high-SES peers for both low-SES and high-SES individuals across the six settings, among those who are assigned to a group in that setting. Low-SES individuals' exposure to high-SES peers is below 1 (i.e., fewer than 50% of their peers have above-median SES) on average in all settings except in colleges, where exposure is above 1 because most people who attend

college have high SES. In contrast, for high-SES individuals, exposure to high-SES peers is well above 1 in all settings. This disparity in exposure reflects segregation across groups; for instance, high-SES people tend to attend different religious institutions and colleges than low-SES people do, as is well known from prior studies on segregation.

Social network data allow us to go beyond measures of segregation and analyze differences in rates of *interaction* conditional on exposure. This ability to identify interaction (friendship) rather than merely exposure (geographical proximity) is a key distinction between the present study and recent work that measures experienced segregation using location data from mobile devices (Athey et al. 2020; Dong et al. 2020; Levy et al. 2020; Reme et al. 2022; Wang et al. 2018). Figure 2c shows mean rates of friending bias—the extent to which rates of *friendship* with high-SES individuals deviate from rates of *exposure* to high-SES individuals—across settings. The green bars show that levels of friending bias for low-SES individuals are typically positive, but differ substantially across settings.

Friending bias is highest on average in neighborhoods, where mean friending bias for low-SES individuals is 0.17. That is, low-SES individuals befriend high-SES people in their ZIP codes at a 17% lower rate than would be the case if they were to befriend high-SES individuals in proportion to their presence in their ZIP codes. Friending bias may be high at the neighborhood level partly because of residential segregation *within* ZIP codes that limits opportunities for contact and interaction between low- and high-SES people.

Friending bias is lowest on average in religious groups, where friending bias is -0.03, implying that low-SES individuals tend to form friendships with high-SES members of their religious groups at a rate that is slightly *higher* than the share of high-SES individuals in their religious groups. Friending bias is negative in religious groups because religious-group friendships do not exhibit substantial homophily by SES—a finding consistent with prior research using survey data (Wuthnow 2002)—and because high-SES people make more friends than low-SES people. Holding fixed exposure, low-SES people are about 20% more likely to befriend a given high-SES person in their religious groups than in their neighborhoods—a large difference, comparable in magnitude to the 22.4% under-representation of high-SES friends on average among low-SES individuals (Chetty et al. 2022). Put differently, if friending bias in all settings was reduced by an amount equal to the difference in friending bias between neighbourhoods and religious groups, most of the disconnection between low-SES and high-SES individuals in the US would be eliminated.

Because religious groups are highly segregated by income, as shown in Figure 2b, their low friending bias does not currently translate to a high level of economic connectedness (see Figure 2a). Efforts to integrate religious groups by SES may be particularly effective at increasing economic connectedness

if friending bias remains low as they became more integrated. This assumption is not innocuous—as illustrated by the challenges faced in efforts to integrate college classrooms (Carrell et al. 2013)—but it is bolstered by the fact that religious groups exhibit low levels of friending bias at all levels of exposure (Supplementary Figure 1b).

The orange bars in Figure 2c show that across all settings, high-SES people are more likely to befriend their fellow high-SES group members (and correspondingly less likely to befriend low-SES group members) than what one would expect based solely on the socioeconomic compositions of their groups. Again, there is sizable heterogeneity in friending bias across settings: high-SES people exhibit the most friending bias (in absolute terms) in neighborhoods, and the least in recreational groups.

A natural question that arises from these differences in friending bias across settings is whether they are an attribute of the setting itself, or a reflection of the types of individuals who join that setting. For example, are religious groups particularly good at fostering ties between low- and high-SES members, or are low-SES individuals who attend religious groups more likely to form such ties across all settings? To distinguish between these explanations, Figure 2d plots friending bias in each of the other five settings minus friending bias in religious groups for low-SES members of religious groups. Members of religious groups exhibit much more friending bias in all other settings than they do in religious groups, showing that the settings where friendships form matter.

The fact that friending bias varies significantly across settings suggests that it is in significant part determined by the nature of the institutions in which people interact—consistent with Blau's theory of social structure (Blau 1977)—rather than entirely determined by preferences. This result suggests that friending bias can potentially be changed through policy interventions (e.g., by changing the structure of the groups in which people interact), much as the socioeconomic composition of groups can be influenced by policy (e.g., via busing or affordable housing programs). How important is it to reduce friending bias vs. increase exposure to increase economic connectedness? We now turn to exploring the relative contribution of these factors in generating differences in connectedness.

III.B Decomposing Differences in Connectedness by SES

We quantify how much of the difference in the share of high-SES friends between low- and high-SES people is due to differences in friending shares across settings, differences in exposure, and differences in friending bias by conducting counterfactual exercises that sequentially remove variation in each of these three dimensions (see Section VI.D for details). Conceptually, our goal is to determine how much of the difference in connectedness would remain if low-SES people made friends in different settings at the same rates as high-SES do (same friending shares); if they participated in groups with the same

shares of high-SES members (same exposure); and if they made high-SES friends at the same rates conditional on exposure as high-SES individuals do (same friending bias).

Figure 3a presents the results of this exercise. The top bar shows that EC for the average low-SES individual is 0.83, while the bottom bar shows that EC is 1.53 for the average high-SES individual—corresponding to a gap in economic connectedness of 0.7. Now consider equating the share of friends that the average low-SES person makes across the six settings to match that of the average high-SES person. Intuitively, this exercise asks: holding constant the exposure and friending bias that low-SES people have across settings, what would happen to low-SES individuals' economic connectedness if they were to make friends at the same relative rates across settings as high-SES people? For example, this counterfactual would increase the overall share of friends low-SES people make in college to match that of high-SES people; however, it would not change the specific colleges that low-SES people attend to match those of high-SES people (as changes in groups within settings would generate a change in exposure).

The second bar in Figure 3a shows that equating friending shares across settings by SES closes only 12% of the gap in EC between the average low- and high-SES person. Hence, differences in the settings where people make friends explain little of why high-SES individuals have more high-SES friends. This is consistent with the fact that the variation in EC across settings for low-SES individuals is small compared to differences in EC by SES within each setting (Figure 2a): even if low-SES individuals were to make all their friends in their highest-EC setting (colleges), their EC would still be substantially below that of high-SES individuals.

Next, we preserve these equated friend shares across settings and set the exposure rates in each setting for the average low-SES person to match the exposure rate for the average high-SES person in that setting. This counterfactual resembles a desegregation policy that adjusts the socioeconomic composition of groups but leaves friendship patterns within them unchanged. For example, in the context of colleges, this counterfactual can be interpreted as having low-SES students attend the same colleges as high-SES students, but retaining their current rate of befriending a given high-SES college peer. The third bar in Figure 3a shows that equating exposure in addition to friending shares would increase the EC of the average low-SES individual to 1.21, closing 54% of the gap in EC between the average low-SES and high-SES person. Intuitively, this is because the gap in exposure by SES in Figure 2b is roughly half as large as the gap in EC in Figure 2a in most groups. While a 54% reduction is substantial, it implies that even if neighborhoods (ZIP codes), schools, colleges, etc., were perfectly integrated by SES, nearly half of the gap in economic connectedness between low- and high-SES individuals would remain.

In the fifth bar, we further set friending bias in each setting for the average low-SES person equal to friending bias of the average high-SES person in that setting. Equating friending bias mechanically closes the remaining 46% gap in EC between the average low- and high-SES person.

III.C Decomposing Differences in Connectedness Across Areas

We use a similar approach to analyze why economic connectedness among low-SES people varies geographically (Chetty et al. 2022). We begin by collapsing our individual-level measures of exposure and bias to the county level, calculating mean high-SES exposure and friending bias among low-SES individuals for each county. Figure 4 maps these variables by county. In addition, we provide an illustrative example of local-area variation by presenting maps of these variables by ZIP code in the Los Angeles metropolitan area. As one might expect, exposure is generally higher in places with higher average incomes (Supplementary Information C.2), such as along each coast of the continental United States and near the coast in the Los Angeles metropolitan area. Friending bias is lowest in the Midwest and Great Plains. Friending bias is lower on average in areas with more high-SES exposure, with a correlation of about -0.2 across counties, but there are many exceptions to this pattern. For example, the Northeast generally has high exposure but also high friending bias (i.e., low- and high-SES people in the Northeast are relatively well integrated in schools and neighborhoods, but tend to befriend each other at lower rates).

We use these area-level statistics to decompose the sources of the ZIP-code-level variation in the economic connectedness of low-SES individuals (see Section VI.D for details). The top bar in Figure 3b shows that the average EC for low-SES people living in ZIP codes that are in the bottom quintile of the national distribution of ZIP-code-level EC averages is 0.52. The bottom bar shows that the corresponding value for low-SES people living in the top quintile of ZIP codes (again in terms of average levels of EC among low-SES individuals) is 1.22. The bars in the middle decompose this top-to-bottom quintile difference in EC by sequentially equating the share of friends made in different settings, rates of exposure to high-SES peers, and rates of friending bias of the average low-SES person in bottom-EC-quintile ZIP codes to match the corresponding values for the average low-SES person in top-EC-quintile ZIP codes (see Supplementary Information B.5 for details). We find that 73% of the difference in EC between ZIP codes in the bottom and top quintiles of the EC distribution is explained by differences in exposure, while 16% is explained by differences in friending bias, and 11% by differences in friending rates across settings.

The geographical variation in EC is driven primarily by differences in exposure because high-SES exposure varies more at a geographical level, whereas friending bias varies more across settings (e.g.,

between neighborhoods and religious groups). The variation in exposure is 3.3 times larger across counties than across settings (Table 1). In contrast, the variation in friending bias is 3.3 times larger across settings than across counties. Intuitively, in areas where the share of high-SES people is high in one setting (e.g., in neighborhoods), it is generally high in other settings as well (e.g., in schools). Conversely, friending bias tends to be relatively consistent by setting across geographies, with low-bias settings in one area (e.g., religious groups) generally exhibiting low friending bias in other areas as well. In short, where one lives influences one's exposure to high-SES individuals, but the groups in which one participates substantially shape the extent to which one interacts with those high-SES peers.

In summary, differences in high-SES exposure generate most of the variation in economic connectedness of low-SES people across *areas*, but friending bias and exposure contribute about equally to explaining the difference in the share of high-SES friends between low- and high-SES *people*. The reason is that exposure varies more across areas than it does by individual socioeconomic status, whereas friending bias differs sharply by SES and is relatively stable (but large) across areas.

III.D Exposure, Friending Bias, and Upward Mobility

Given that both exposure and friending bias contribute to differences in economic connectedness, we next examine whether the strong correlation between economic connectedness and upward income mobility documented in the previous paper in this series (Chetty et al. 2022) is driven by one or both of these components. We define upward mobility as the average income rank in adulthood of children who grew up in families at the 25th percentile of the national income distribution in a given county or ZIP code, drawing on data from the Opportunity Atlas (Chetty et al. 2018).

In column 1 of Table 2, we regress log upward mobility on log economic connectedness across ZIP codes (see Section VI.E for details). We find an elasticity of upward mobility with respect to EC of 0.24: a 10% increase in EC is associated with a 2.4% increase in upward mobility. In column 2, we regress log upward mobility on log exposure and log(1-friending bias). We find strong associations between both exposure and friending bias and measures of upward mobility, with elasticities of 0.25 and 0.19, respectively. Next, we examine how these relationships vary within vs. across counties. Columns 3 and 4 of Table 2 include county fixed effects in the specifications from columns 1 and 2. When comparing ZIP codes within counties, higher exposure and lower friending bias remain strongly associated with higher levels of economic mobility, with elasticities just under 0.25. In columns 5 and 6, we conversely focus on across-county variation by replicating columns 1 and 2 at the county level. We find qualitatively similar effects, though the estimates of the effects of friending bias on economic mobility become less precise, largely because most of the variation in friending bias is within rather than across counties (Table 1).

In column 7 of Table 2, we change the dependent variable in the regression to the log of each county's *causal effect* on upward mobility as estimated by Chetty and Hendren (2018) based on analyzing movers (see our companion paper Chetty et al. (2022) for further details on the interpretation of these causal effect measures). Both exposure and friending bias remain strongly predictive of counties' causal effects on upward mobility, implying that moving to a place with greater exposure or lower friending bias at an earlier age increases the earnings in adulthood of children who grow up in low-income families.

We conclude that the relationship between economic connectedness and upward mobility is not driven merely by the presence of high-SES peers (e.g., through the availability of additional resources for schools financed by local property taxes). Instead, *interaction* with those peers is what predicts upward mobility most strongly (see Supplementary Information C.3 for further discussion). In the context of schools, this result implies that the average income of classmates predicts upward mobility for children with low SES insofar as it affects the extent of their social interactions with high-SES students. Combined with our finding that friending bias accounts for roughly half the difference in the share of high-SES friends between low- and high-SES people, these results imply that increasing economic connectedness—the form of social capital most strongly associated with economic mobility—would require efforts to both increase integration (exposure) and reduce friending bias within groups. In the next section, we show how our data can inform which of these approaches is likely to be most effective in a given group.

IV Targeting Interventions to Increase Economic Connectedness

IV.A Exposure and Friending Bias by High School

Having shown how exposure and friending bias vary across settings and areas, we now analyze variation in these statistics across the groups that comprise a given setting (e.g., each high school in the "high school" category). We begin by examining variation across high schools and then turn to variation across colleges. We publicly release estimates of exposure and friending bias for each high school and college as well as by neighborhood (ZIP code); for religious organizations, recreational groups, and employers, sample sizes are too small to obtain reliable estimates at the group-specific level.

For high schools, we report estimates based both on students' own (post-high-school) SES in adulthood—the same SES measure that was analyzed above—as well as estimates based on their parents' SES (see Section VI.F). These measures have different applications. Measures of economic connectedness based on parental SES are relevant for policy discussions at the school level, which often focus on the degree of connection between children from different parental backgrounds. Measures based on own SES are useful for understanding the environments in which friendships between low-SES and

high-SES adults are formed, i.e., the extent to which a school might contribute to levels of economic connectedness in the next generation. Although the two measures capture different concepts, they yield fairly similar rankings of schools in terms of exposure and friending bias: the correlation between the two measures is 0.84 for exposure and 0.59 for friending bias across schools (Supplementary Table 1). We therefore focus on the parental SES measure here and present analogous results using own SES in Supplementary Figure 2.

Figure 5a plots friending bias (with an inverted vertical scale, so that moving up correponds to *less* bias) vs. the share of high-parental-SES students (i.e., half of high-SES exposure) by high school. Both exposure to students with high-SES parents (i.e., socioeconomic composition) and friending bias vary substantially across schools. The reliability of the exposure estimates, estimated using a split-sample approach (see Section VI.F), is 0.99 at the school level; that is, 99% of the variance in exposure reflects true differences in the share of students with high-SES parents rather than sampling error. The reliability of the friending bias estimates is 0.58. This implies that a school that we estimate to have a 10% higher friending bias estimate will, on average, exhibit 5.8% higher bias in future cohorts. Estimates of exposure and friending bias based on own SES have higher reliabilities (0.99 for exposure and 0.88 for friending bias) because they use the full sample rather than just the subset of individuals we can link to their parents.

Friending bias differs considerably even among nearby schools with similar socioeconomic compositions. For example, Walter Payton College Preparatory High School and Evanston Township High School (ETHS) are two large high schools in the Chicago metro area that have similar fractions of students from above-median-SES families. However, ETHS has much higher friending bias than Payton: low-SES students at ETHS are much less likely to befriend their high-SES peers than low-SES students at Payton are, consistent with prior ethnographic evidence documenting high levels of friending bias at ETHS (Supplementary Information C.4). One potential explanation for this difference is the greater similarity of students on other dimensions at Payton relative to ETHS. Payton is a public magnet school that requires an entrance exam of all students. In contrast, ETHS is open to all students residing in the local catchment area, resulting in a more heterogeneous student body in terms of academic preparation—and concomitant segregation of classes—that may lead to higher friending bias (Villano 2011).

Predictors of Friending Bias

Building on this comparison, we examine the factors that predict friending bias across high schools more systematically by correlating bias across schools with various observable characteristics. Consistent with the ETHS-Payton comparison, we find that friending bias is higher on average in schools with more

academic tracking as measured by enrollment rates in gifted and talented or Advanced Placement classes (Figures 6a and 6b). Friending bias is generally lower in smaller schools (Figure 6c), consistent with previous work documenting less homophily in smaller groups (Blau 1977; Cheng and Xie 2013; Currarini et al. 2010).

In Figure 6d, we examine the relationship between friending bias and a school's share of high-parental-SES students. This relationship is non-monotonic, with friending bias highest in schools with a roughly equal representation of students from below- and above-median SES families. This may be because there is less scope for low-SES or high-SES students to develop homogeneous cliques when there are relatively few members of their own group. Friending bias is also higher in more racially diverse schools, as measured by a Herfindahl-Hirschman index (Figure 6e) or the share of white students in the school (Figure 6f). One potential explanation for the link between racial diversity and friending bias by SES is that when low- and high-SES students have different racial and ethnic backgrounds, they are less likely to be friends.

There are similar associations between these factors and friending bias between students who go on to have different socioeconomic statuses in adulthood themselves (Figure 7). In particular, in smaller and less racially diverse schools, there are more friendships between students who go on to have low and high SES in adulthood. We also find similar relationships between friending bias and group characteristics in other settings: higher levels of friending bias are associated with greater racial diversity across colleges and neighborhoods (Figure 8) and larger group sizes across all six settings (Supplementary Figure 1).

The explanatory factors considered in Figures 6-8 are not intended to be exhaustive, and much more remains to be learned about the determinants of friending bias. The main lesson we draw from these correlations is that, much like exposure, friending bias appears to be related to structural factors that can potentially be changed by policy interventions, such as reducing the size of groups and redesigning the nature of academic tracking within schools.

Increasing Connectedness

The variation in exposure and friending bias across schools documented in Figure 5a implies that the most effective approach to increasing economic connectedness differs across schools. To increase economic connectedness in schools in the bottom half of Figure 5a—such as Evanston Township High School, Berkeley High in Berkeley, CA, or Lake Highlands High in Lake Highlands, TX—decreasing friending bias (i.e., increasing social interaction between students from different backgrounds) is likely to be valuable. For example, reducing friending bias at ETHS to zero would result in an increase of 0.15 (15 percentage points) in economic connectedness (measured in terms of parental SES). To benchmark this impact, note that the average parental EC among high school friends of low-parental-

SES individuals across the schools in Figure 5a is 0.92. This implies that in the average U.S. high school, students with low-SES parents have 8% fewer high-parental-SES friends than one would expect in a scenario where students with high and low SES made the same total number of high school friends and exhibit no homophily. The current level of friending bias at ETHS thus reduces the share of high-SES friends among low-SES students by almost twice the degree of under-representation of high-SES friends among students from low-SES families at the average U.S. high school (15% vs. 8%). Hence, at schools like ETHS, increasing cross-SES interaction within the student body may be a more effective way to increase economic connectedness than attempting to further diversify the student body. In contrast, for schools that exhibit low levels of exposure and low levels of bias, such as West Charlotte High or LeFlore Magnet (shown in the top left quadrant of Figure 5a), increasing socioeconomic integration (exposure) is a necessary first step to increasing economic connectedness.

The preceding analysis focuses on how to maximize economic connectedness from the perspective of a *given* low-SES student (i.e., how to increase the likelihood that they form cross-SES friendships within a given school). However, from a social perspective, it may be more relevant to consider a given school's contribution to the total number of cross-SES friendships in society. To see how these concepts differ, consider Phillips Exeter Academy, an elite private boarding school in New Hampshire where more than 80% of students come from above-median-SES families (exposure is high) and friending bias is low (below zero). Given these conditions, low-SES Phillips Exeter students tend to form many friendships with their high-SES classmates and have high economic connectedness. However, because low-SES students make up only a small share of Phillips Exeter's students, the total number of cross-SES connections that Phillips Exeter generates is relatively small. If Phillips Exeter were to enroll more low-SES (and fewer high-SES) students, it could increase its total contribution to connectedness despite reducing economic connectedness for current low-SES students (as they would be exposed to fewer high-SES peers).

We measure each school's *total contribution to economic connectedness* (TCEC) as the product of the share of low-SES students and the average economic connectedness among low-SES students in that school (see Section VI.G for an algebraic definition). TCEC measures how many friendships a school creates between high- and low-SES students, holding fixed total enrollment and the total number of friends that students make across schools. Reducing friending bias at a school (all else equal) always increases the total number of friendships between low- and high-SES students. However, increasing the share of high-SES students has non-monotonic effects on TCEC. Schools that have very few high-SES students offer few opportunities for their low-SES students to meet high-SES peers and thus contribute little to overall economic connectedness. Conversely, schools that have predominantly

high-SES students, such as Phillips Exeter or the Dalton School in New York City, provide many high-SES connections to the low-SES students that they *do* enroll, but offer those opportunities to relatively few low-SES students and thus also have low $TCEC$.

Because of these competing forces, holding friending bias fixed, an above-median-SES share of 50% (i.e., achieving perfect socioeconomic integration) maximizes the total number of cross-class connections at a school. Schools that have low friending bias and near-equal representation of students whose parents have below- and above-median SES—such as Lane Technical in Figure 5a—contribute the most to total economic connectedness in an accounting sense. More generally, the direction in which one must shift exposure to increase the total number of cross-SES links differs based on a school's initial share of high-SES students. In contrast, reducing friending bias always increases EC for a given low-SES student as well TCEC.

Furthermore, increasing the share of high-SES students in one school necessarily requires reducing the share of high-SES students in at least one other school, since the total number of above-median-SES students is fixed. As a result, increasing high-SES shares even at schools where high-SES shares are below 50% can have ambiguous effects on economic connectedness in society as a whole. If the high-SES students who join a given school A otherwise would have attended school B where they would have connected with more low-SES peers, overall EC in society could fall even though $TCEC$ at school A would rise. Hence, one must be cognizant of the counterfactual distribution of SES across schools when evaluating the impacts of increasing exposure. Efforts to reduce friending bias in a given school, by contrast, do not generally have direct implications for connectedness at other schools.

In sum, for schools that already have diverse student bodies (i.e., schools that have a balanced socioeconomic representation) but high levels of friending bias, initiatives to identify and address institutional factors contributing to friending bias may be the most fruitful path to increasing their total contributions to connectedness. For schools that currently have less diverse student bodies, it may be valuable to increase diversity in a manner that takes account of which schools new students would otherwise have attended.

IV.B Exposure and Friending Bias by College

Figure 5b replicates Figure 5a for colleges, again using parental SES. We see analogous heterogeneity in exposure and friending bias across colleges, with similar implications. For example, Yale University exhibits relatively low friending bias and has a large high-SES share, resulting in high levels of economic connectedness for its low-SES students. However, because low-SES students only make up a small share of the student body, Yale, like many other elite private colleges, creates relatively few cross-SES

connections (it has low TCEC).

Among colleges with more socioeconomically diverse student bodies, such as Wayne State and Howard, there is again considerable variation in connectedness that results from differences in friending bias. Similar to high schools, friending bias tends to increase with a college's size and with the degree of racial diversity of the student body (Supplementary Figure 3). In a different vein, ethnographic evidence suggests that many universities that exhibit high levels of bias—such as the University of Alabama, Syracuse University, or the University of Mississippi—feature significant Greek life, where high costs of fraternity and sorority dues may generate friending bias on campus (Terry 2014). Similarly, community colleges without a substantial residential student population (e.g., the City College of San Francisco or San Diego City College) tend to exhibit high levels of friending bias. Systematically evaluating these and other hypotheses using the data constructed here would be a useful direction for further work. For now, these results again suggest that friending bias is at least partly determined by structural factors that could potentially be changed by colleges, much like recent efforts to increase socioeconomic diversity at elite private colleges.

IV.C Causal Effects of Integration on Connectedness

Having established that there is significant variation across schools and colleges in friending bias and exposure, we now examine whether these estimates are sufficiently reliable to determine what interventions will be most effective at increasing economic connectedness in a given school. As a practical illustration, consider policies that seek to increase socioeconomic diversity in a given school district. Can our estimates of average friending bias be used to reliably identify the schools where such policies will increase connectedness the most? If estimates of friending bias are perfectly stable, the impact of a change in socioeconomic composition will be well predicted by historical estimates of average friending bias. In contrast, if estimates of bias change over time (e.g., due to measurement error or drift), or if the effects of incremental changes in socioeconomic diversity differ substantially from historical averages of friending bias, predictions based on existing observational data may not provide reliable forecasts. It is thus an empirical question whether the school-level estimates we report provide useful information to predict the impacts of policy changes. We use two quasi-experimental research designs to identify the causal effects of changes in exposure on connectedness—cross-cohort fluctuations and regression discontinuity—and show that our school-level estimates of average friending bias predict the causal effects of these changes in exposure on economic connectedness.

Cross-Cohort Fluctuations

In our first approach, we analyze the effects of fluctuations in the share of high-SES students across

cohorts within a high school on students' friendship patterns. Such fluctuations in cohort composition are largely a consequence of random variation in the student body, as discussed in Section VI.H. Intuitively, we compare low-SES students who attend the same school and examine whether those who happen to be in cohorts that have a larger share of high-SES students tend to have more high-SES friends as a result. To harness more variation across cohorts, we focus here on connections between individuals with parents in the lowest and highest SES *quintiles* (rather than below- vs. above-median SES, as we do in the rest of the paper).

Figure 9a presents a binned scatter plot of changes in economic connectedness for low-SES students across cohorts within a school vs. cross-cohort changes in high-SES exposure (see Section VI.H for details). In this analysis, we focus on measuring within-cohort economic connectedness and exposure—i.e., the share of high-SES friends and high-SES peers that low-SES students have within only their own cohorts in their high schools. The strong positive relationship demonstrates that, within a given school, students in cohorts that happen to have more high-SES students have significantly more high-SES friends in their cohorts on average. Hence, greater high-SES exposure translates to significantly greater high-SES friendships on average, showing that socioeconomic integration can be a powerful tool for increasing cross-class interaction.

The slope of 0.89 in Figure 9a implies marginal friending bias of 0.11: a 10 percentage point increase in the share of high-SES peers in a given cohort leads to an 8.9 percentage point increase in the share of high-SES friends among low-SES students in that cohort on average. The corresponding cross-sectional mean of bottom-to-top-parental-SES-quintile friending bias is also 11%. Hence, an *incremental* change in socioeconomic integration has a similar causal impact on connectedness to what one would predict based on the *average* level of friending bias in the observational data.

Next, consider how the relationship in Figure 9a varies across schools that have different levels of friending bias. We estimate a regression analogous to that shown in Figure 9a separately for school-cohort cells in each decile of the friending bias distribution (estimating friending bias based on data for other cohorts in the same school). Figure 9b plots the estimated regression coefficients in each decile against the level of friending bias in that decile. There is a strong negative relationship, showing that an increase in high-SES exposure produces fewer cross-SES friendships in schools that exhibit higher friending bias. The slope of the relationship in Figure 9b is -0.61, implying that a 1 percentage point increase in mean friending bias in other cohorts translates to a 0.61 percentage point reduction in the impact of exposure on EC.

This coefficient may be below 1 for two different reasons. First, sampling error in our estimates of friending bias leads to imperfect predictions of friending bias in a given cohort. Second, the average level

of friending bias observed in a school may not correspond to the bias associated with befriending an incremental high-SES student in a cohort. To distinguish between these explanations, note that in the sample used for the quasi-experimental analysis in Figure 9, a 1 percent increase in mean friending bias in other cohorts is associated with a 0.67 pp increase in friending bias in a given cohort c on average. Correcting for this degree of attenuation bias, the implied impact of a 1% increase in average friending bias in a given cohort is a $0.61/0.67 = 0.91$ percentage point reduction in the impact of an incremental change in exposure on EC. Hence, fluctuations in exposure translate to cross-SES friendships at close to the rate one would expect given the average friending bias in a given cohort. This finding support the use of average observed friending bias in a school to predict the effects of incremental changes in exposure on EC, in particular after accounting for sampling error in friending bias.

Regression Discontinuity

If high-SES students move into certain school districts over time and those districts also exhibit secular trends in cross-SES friendships (e.g., due to changes in friending bias) for other unrelated reasons, the cross-cohort comparisons above may yield biased estimates of the causal effect of exposure on EC. To address such concerns, we now turn to a second approach that leverages the fact that most states use cutoffs based on birth dates to determine when students begin school; for example, in Texas, students who turn five on or before September 1 begin Kindergarten that year, while those who turn five on or after September 2 begin school the next year (Supplementary Table 2). We use these cutoffs to implement a regression discontinuity design, comparing economic connectedness for low-SES individuals who happen to fall on different sides of the entry cutoff (e.g., are born on September 1 vs. September 2) and are thus exposed to high school peer groups that differ in their share of high-SES students. See Section VI.I for a discussion of the identification assumptions underlying this design and further details.

We begin by focusing on pairs of adjacent cohorts where the magnitude of the jump in the share of high-parental-SES students is large, i.e., lies in the top quartile of the distribution of changes in high-SES shares. In Figure 10a, we examine how these jumps in exposure to high-parental-SES peers affect within-cohort economic connectedness. We examine these effects separately in schools with low (bottom quartile) vs. high (top quartile) friending bias. The share of high-parental-SES friends jumps to the right of the school entry cutoff in both sets of schools, showing that exposure to more high-SES peers in one's cohort (i.e., greater exposure) leads students to form more high-SES friendships within their school cohorts. However, the magnitude of the jump in high-SES friendships caused by this increased exposure is 0.06 units larger in low-friending-bias schools than in high-friending-bias schools. This difference is similar to the observed difference in average friending bias between schools classified (based on data from other cohorts) to be in the bottom vs. top quartile of friending bias,

again demonstrating that observed average friending bias (adjusted for measurement error) predicts the effect of incremental changes in exposure on economic connectedness accurately.

In Figure 10b, we extend this approach to look beyond cohort pairs with large fluctuations in high-SES shares. We plot regression-discontinuity estimates of changes in within-cohort EC vs. changes in exposure for each of the four quartiles of changes in high-SES exposure across cohorts, separately for schools in the bottom and top friending bias quartiles. The right-most points in this figure match the RD estimates reported in Figure 10a. Low-SES students' shares of high-SES friends increase linearly with their exposure to high-SES peers across the distribution of exposure changes. The slope of the line is steeper in schools with low friending bias, showing that greater high-SES exposure translates to greater cross-SES friendships when friending bias is low.

We conclude that our school-specific observational estimates of friending bias are sufficiently stable and reliable for predicting the causal effects of changes in exposure on economic connectedness out of sample, and hence can inform where efforts to reduce friending bias vs. increase exposure are likely to be most valuable.

V Discussion

The extent to which individuals interact across class lines depends on both exposure (the socioeconomic composition of the groups to which people belong) and friending bias (the rate at which cross-SES friendships are formed conditional on exposure). To date, there have been extensive policy efforts on the exposure dimension, such as busing programs aimed at integrating schools; zoning and affordable housing policies aimed at integrating neighborhoods; and college admissions reforms to boost diversity on campuses. Such interventions to increase integration can increase cross-SES interaction substantially. However, even if all such groups were perfectly integrated by socioeconomic status, half of the social disconnection between low- and high-SES people would persist because of friending bias within groups.

Our analysis suggests that friending bias, like exposure, is shaped by social structures and institutions and can therefore be influenced by policy changes. Although interventions to reduce friending bias have been studied less extensively (Alan et al. 2020; Pallais and Anders 2021; Rohrer et al. 2021), there are several recent initiatives that seek to reduce friending bias. We conclude by discussing some examples to illustrate the types of interventions that could be studied and scaled going forward.

(1) *Changes in Group Size and Tracking.* As discussed above, Berkeley High School (BHS) has historically been socioeconomically diverse, but has had high levels of friending bias. Those familiar with the school were aware of this issue and pointed to tracking as a source of substantial within-school segregation. For example, Kim (2018) writes: "BHS's population of more than 3,000 students is cur-

rently split into five learning communities, each meant to provide its own focus and curriculum....causing implicit segregation, resulting in student learning communities with separated concentrations of white students and students of color." In an attempt to overcome this within-school segregation and to reduce the associated friending bias, in 2018 BHS began assigning students to small, intentionally diverse "houses" or "hives" in ninth grade. Such attention to the way in which students are tracked to different classes within schools and the size of the groups in which they participate outside class may be helpful in reducing friending bias more broadly.

(2) *Restructuring of Space and Urban Planning.* Lake Highlands High School in Texas is another school in which we observed high levels of friending bias (Figure 5a). In this case, administrators and students at Lake Highlands High identified the architecture of the building as an impediment to cross-SES interaction: "At Lake Highlands High, the duplicated rooms—cafeterias, libraries, science labs—led to unintentional student segregation," such that "students clustered in one of three lunchrooms depending on their social group or the options for low-cost and free lunch" (HKS 2019). The school recently attempted to reduce this source of friending bias through a large-scale construction project that created a single cafeteria and more spaces for all students to interact. One of the architects described the project's goals as follows: "shrink income-based inequalities in education by designing schools that improve the way students learn and socialize," noting that "though students may still split into their own cliques [...], they'll have more opportunities to cross paths and interact with peers from other social groups." Architecture and urban planning could play a role in reducing friending bias outside schools as well. Examples include "social infrastructure" such as public libraries to build social bonds across groups (Klinenberg 2018); the effects of public parks on social interactions (Abbiasov 2021); and the impacts of public transit on interactions between people living in different neighborhoods (Bailey et al. 2020a).

(3) *New Domains for Interaction.* Another approach to reducing friending bias could be to create new programs and venues for cross-SES interaction. For example, the Boston gym Inner City Weightlifting (ICW) began a program to increase cross-SES connections by recruiting personal trainers from lower-SES backgrounds to coach their more affluent clients. Founder Jonathan Feinman describes the objective as follows: "At ICW, through our career track in personal training, we help create economic mobility for people in our program as they begin earning $20-$60 per hour training clients from opposite socioeconomic backgrounds. More importantly, this flips power dynamics, bridges social capital, and creates a genuine form of inclusion that disrupts the system of segregation, isolation, and racism that leads to the streets. The people in our program gain access to new networks and opportunities, while our clients gain new insights and perspectives into complex social challenges" (Feinman 2022). Feinman

notes that the program appears to be a success: "Along the way, something unexpected happened. We had our paying clients—people paying our student trainers—visiting our students in jail when things went wrong. They were showing up in court to be a support. They started offering job opportunities to our students outside of the gym, and they paid for the children of our students to go to summer camp with their own children" (Boston Voyager 2018). More generally, creating new programs and venues for cross-SES interaction (e.g., through peer mentoring programs or internship programs) could help reduce friending bias.

The school-, college-, and ZIP-code-level data on exposure and friending bias made publicly available through this project can help to determine whether interventions to reduce friending bias or efforts to increase socioeconomic diversity are likely to be most valuable for increasing economic connectedness. Going forward, the methods and data developed here can be used to evaluate the causal effects of interventions such as those described above. By studying changes in exposure and friending bias over time, researchers and policy makers can learn from places where progress is being made and provide assistance to communities seeking to improve economic connectedness—the form of social capital most strongly associated with economic mobility.

VI Methods

VI.A Sample Construction

This section describes the methods used to generate the data analyzed in this paper. A server-side analysis script was designed to automatically process the raw data, strip the data of personal identifiers, and generate aggregate results, which we analyzed to produce the conclusions in this paper. The script then promptly deleted the raw data generated for this project (see Section VI.J).

We start from the analysis sample constructed in Chetty et al. (2022): users aged between 25 and 44 as of May 28, 2022, who reside in the United States, were active on the Facebook platform at least once in the prior 30 days, have at least 100 U.S.-based Facebook friends, and have a non-missing predicted ZIP code. We then restrict attention to individuals for whom we can allocate at least one friendship to the group in which it was formed (using the approach described below). The resulting sample consists of 70.3 million Facebook users, corresponding to 82% of the U.S. population between ages 25–44 based on the American Community Survey (ACS).

We do not link any external individual-level information to the Facebook data. However, the project uses various publicly available sources of aggregate statistics to supplement the analysis, such as data on median incomes by block group from the 2014–18 ACS; school-level variables from the National Center for Education Statistics (NCES) and the Civil Rights Data Collection (CRDC); and various

college-level statistics from the Integrated Postsecondary Education Data System (IPEDS) and Chetty et al. (2022). We describe those data in detail in Supplementary Information A.

VI.B Variable Definitions

We construct the following sets of variables for each person in our analysis sample; the first four variables are identical to those used in Chetty et al. (2022), while the fifth is new to this paper.

Friendship Links. The data contain information on all friendship links between Facebook users. Facebook friendship links need to be confirmed by both parties, and most links are between individuals who have interacted in person (Jones et al. 2013). As a result, the Facebook friendship network can be interpreted as providing data on people's "real-world" friends and acquaintances rather than purely online connections.

Locations. Every individual in our analysis sample is assigned a ZIP code and county based on information and activity on Facebook, including the city stated on their Facebook profile as well as device and connection information. Formally, we use 2010 Census ZIP Code Tabulation Areas (ZCTAs) to perform all geographic analyses of ZIP code-level data. We refer to these ZCTAs as ZIP codes for simplicity. According to the 2014–2018 ACS, there are 219,214 Census block groups, 32,799 ZIP codes, and 3,220 counties, with average populations of 1,488, 9,948, and 101,332 in each respective geographic designation.

Socioeconomic Status. Social scientists have measured "socioeconomic status" using many different variables, ranging from income and wealth to educational attainment, occupation, family background, neighborhood, and consumption (Karl R White 1982a). To capture these varied definitions, we construct a model that generates a composite measure of SES for working-age adults (individuals between the ages of 25 and 64) that combines various characteristics (see Section VI.J for a discussion of how user privacy was protected during this project). We construct our baseline SES measure in three steps (see Supplementary Information B.1 of Chetty et al. 2022 for details).

First, for Facebook users who have Location History (LH) settings enabled, we compute the median household income in their Census block groups. Location History is an opt-in setting for Facebook accounts that allows the collection and storage of location signals provided by a device's operating system while the app is running. We observe Census block groups from individuals in the LH subsample; in contrast, we can only assign individuals who do not have LH enabled to ZIP codes. If an individual subsequently opts out of Location History, their previously stored location signals are not retained.

Second, we estimate a gradient-boosted regression tree to predict these median block group household incomes using variables observed for all individuals in our sample, such as age, gender, language,

relationship status, location information (ZIP code), college, donations, phone model price and mobile carrier, usage of Facebook on the web (rather than a mobile device), and other variables related to Facebook usage. We use this model to generate SES predictions for all individuals in our sample.

Finally, individuals (including the LH users in the training sample) are assigned percentile ranks in the national SES distribution based on their predicted SES relative to others in the same birth cohort.

We do not use any information from an individual's friends to predict their SES, ensuring that errors in the SES predictions are not correlated across friends, which would bias our estimates of homophily by SES. We also do not use direct information on individuals' incomes or wealth, as we do not observe these variables at the individual level in our data; however, we show below that our measures of SES are highly correlated with measures of income across subgroups. Note that the algorithm described above is one of many potential ways of combining a set of underlying proxies for SES into a single measure; other methods discussed in Chetty et al. (2022) yield very similar results.

Parental Socioeconomic Status. We link individuals in our primary analysis sample (i.e., those aged 25-44) to their parents (who may not be in the analysis sample themselves) to construct measures of family socioeconomic status during childhood. To link individuals to their parents, we use self-reported familial ties, a hash of user last names, and public user-generated wall posts and major life events (see Chetty et al. 2022 for details). We then use the SES of parents, constructed via the algorithm described above, to assign parental SES to individuals. We are able to assign parental SES ranks for 31% of the primary analysis sample.

Groups Where Friends are Made. We assign friendships to the groups in which they were made by focusing on six *settings* (group types) that we can identify reliably in our data: high schools, colleges, employers, neighborhoods (ZIP codes), faith-based (religious) groups, and recreational groups. These settings span the most common places where users make friends, excluding family (Jeffres et al. 2009; Oldenburg 1999).

We first use self-reported data (for colleges, employers, and high schools), liked pages of places of worship (religious groups), and group membership (recreational and religious groups) to assign individuals to at most one group in each setting (see Supplementary Information B.1 for details). For some people who do not report a high school, we use data on their friendship networks to impute their high school. For the small set of individuals who are members of multiple groups within a setting (e.g., 3.3% of users report multiple high schools conditional on being assigned a high school), we select the group in which the user has the largest number of friends. The quality of our group assignments appears to be high based on comparisons to external statistics. For example, our estimates of the share of high-SES households in each ZIP code, high school, and college have correlations above 0.85 with corresponding

statistics drawn from publicly available administrative data sets (Table 3).

We then assign friendship links to groups based on shared group membership, as described in Supplementary Information B.2. For example, if an individual and one of their friends are part of the same neighborhood, they are identified as neighborhood friends. In cases with shared group membership across multiple settings—for example, when two friends are members of both the same recreational group and workplace—the friendship link is counted in all relevant settings. We are able to allocate about 30% of friendship links to at least one setting. The remaining friends either could not be connected to a group due to missing data (e.g., missing data on the users' or friends' workplace) or were made outside the settings we consider. Note that this research did not involve inferences about an individual's religion; instead, it is focused on whether friendships were formed in a faith-based (religious) group.

VI.C Benchmarking

Table 4a shows summary statistics for the primary analysis sample used in this paper (as of May 28, 2022) and, for comparison, for those between ages 25-44 in the 2014-18 ACS. As discussed in Chetty et al. (2022), the Facebook sample is quite similar to the full population in terms of age, gender, and language. Chetty et al. (2022) further demonstrate that the Facebook sample is broadly representative of the U.S. population geographically and that the SES measures used in our analysis below are well correlated with publicly available statistics and yield estimates of homophily by SES and intergenerational mobility that match external estimates from nationally representative data sets.

When analyzing interventions to increase economic connectedness in high schools and colleges, we focus on the subsample of individuals who can be assigned a high school or college and who can be linked to parents with an SES prediction (in order to measure connectedness by parental SES). Table 4b presents summary statistics for the subsample of 19.4 million who can be assigned parental SES and high school, who constitute 28% of the full analysis sample. The characteristics of this subsample are broadly similar to those of the full sample.

In this paper, we focus on the 30% of friendships that can be assigned to groups in which people interact, which is necessary to identify exposure and friending bias. We find that within the subsample of friendships that can be assigned to groups, homophily is similar to that observed in the full sample of friendships (Figure 11). Moreover, at the individual level, a person's share of high-SES friends in the subsample of friends assigned to a group has a correlation of more than 0.90 with their share of high-SES friends overall. Furthermore, to address potential concerns about bias from under-reporting of groups, we develop a procedure to correct for underreporting of group memberships using external statistics on group membership rates (see Supplementary Information B.3). In this expanded sample,

which accounts for 44% of friendships, our conclusions remain similar (Supplementary Figure 4).

Based on this analysis, we conclude that the subsample of friendships we analyze here is reasonably representative of the broader set of friendships people make on Facebook and in the population in general.

VI.D Decomposing Economic Connectedness

Following Chetty et al. (2022), we define individuals' *economic connectedness* as the extent to which they are friends with high-SES individuals. Formally, let $f_{H,i}$ denote individual i's share of high-SES friends and let $w_H = 0.5$ represent the share of above-median-SES individuals in the population. We define person i's *individual economic connectedness (IEC)* to high-SES individuals as:

$$IEC_{H,i} \equiv \frac{f_{H,i}}{w_H}. \tag{1}$$

If $IEC_{H,i} > 1$, individual i has more high-SES friends than one would expect if friendships were made at random and low- and high-SES people made an equal number of friends; conversely, $IEC_{H,i} < 1$ means that i has fewer high-SES friends than one would expect under random friending.

To decompose $IEC_{H,i}$ into exposure and friending bias, let $\phi_{i,g}$ denote the fraction of friends that individual i makes in group g (out of all friends of i that can be assigned to groups) and G denote the set of all available groups, so that $\sum_{g \in G} \phi_{i,g} = 1$ for each individual. Here, a group g represents a specific school, college, recreational group, etc. to which an individual can belong. Individuals' friending shares $\phi_{i,g}$ are positive or 0 in the specific groups to which they belong and are 0 for all other groups. Let $w_{H,g}$ denote the fraction of members of group g who have high SES and $f_{H,i,g}$ the fraction of friends individual i makes in group g who have high SES (see Supplementary Information B.4 for discussion of how we define $f_{H,i,g}$ when $\phi_{i,g} = 0$).

We can express each individual's connectedness to high-SES individuals as the product of three components, summed across groups:

$$\begin{aligned} IEC_{H,i} = \frac{f_{H,i}}{w_H} &= \sum_{g \in G} \left[\phi_{i,g} \times \frac{f_{H,i,g}}{w_H} \right] = \sum_{g \in G} \left[\phi_{i,g} \times \frac{w_{H,g}}{w_H} \times \frac{f_{H,i,g}}{w_{H,g}} \right] \\ &= \sum_{g \in G} [\phi_{i,g} \times \text{Exposure}_{H,g} \times (1 - \text{Friending Bias}_{H,i,g})], \end{aligned} \tag{2}$$

where

$$\text{Exposure}_{H,g} \equiv \frac{w_{H,g}}{w_H} \tag{3}$$

is the normalized fraction of high-SES individuals in group g. Exposure is below 1 for groups that have a below-average share of high-SES individuals and above 1 for groups that have an above-average share

of high-SES individuals. The final term,

$$\text{Friending Bias}_{H,i,g} \equiv 1 - \frac{f_{H,i,g}}{w_{H,g}}, \tag{4}$$

measures the *deviation* from uniformly random friending conditional on exposure.

If friendships were formed at random and if high- and low-SES people made the same number of friends, then $f_{H,i,g} = w_{H,g}$ and friending bias would be equal to 0. In practice, high-SES individuals make 25.4% more friends than low-SES individuals do on average (Chetty et al. 2022). Maintaining this difference in the number of total friends, in a setting with no homophily by SES (i.e., a setting where high-SES and low-SES individuals have the same probability of befriending a given high-SES person), friending bias would be negative. In particular, if high-SES individuals have $x_g > 1$ times as many friends as low-SES individuals in group g but there is no homophily by SES,

$$\text{Friending Bias}_{H,i,g} = \frac{1 - x_g}{1 + \frac{w_{H,g}}{1-w_{H,g}} x_g} < 0. \tag{5}$$

In a group that is representative of the overall population—where $x_g = 1.254$ and $w_{H,g} = 0.5$—friending bias would be -0.11, and low-SES individuals would have 11% more high-SES friends than the share of high-SES individuals in the group. Relative to this benchmark, a positive value of friending bias constitutes a substantial departure from a society that does not exhibit homophily by SES, as it means that fewer than half of low-SES people's friends are high-SES even though high-SES individuals form more friendships on average.

Decompositions by SES

In the Decomposing Connectedness by SES section, we quantify how much of the difference in the share of high-SES friends between low- and high-SES people is due to differences in friending shares across settings, differences in exposure, and differences in friending bias. Because the parameters in equation 2 vary across individuals and groups even at a given SES level, we take a representative agent approach to decompose the relative contributions of these three factors *on average*. In particular, we consider a representative low-SES agent and high-SES agent who have setting-level friending shares, exposure rates, and friending bias levels that match the means for low- and high-SES people in the population, respectively (see Figure 1 and 2 and Table 5). To conduct the decomposition, we first calculate IECs for the representative low-SES and high-SES agents using these mean values and a formula analogous to equation 2 (see Supplementary Information B.5 for details). We then sequentially set each of the parameters for the low-SES agent to match the values for the high-SES agent, allowing us to determine how much of the difference in the representative-agent IEC across the two SES groups is explained by each of the three factors. We refer to the representative low-SES and high-SES agents as the "average" low- and high-SES persons in the main text.

Because equation 2 is not additive, the share of the total difference attributed to friending bias vs. exposure depends on the order in which we conduct each of the steps of the decomposition exercise. For the decomposition by SES discussed in the main text, we equated rates of exposure before rates of friending bias, effectively examining what the effects of socioeconomic integration would be absent any changes in friending bias. If we instead equate rates of friending bias before rates of exposure (the fourth bar in Figure 3a), 34% of the EC gap would be explained by friending bias and 54% by exposure. Lower friending bias and greater exposure are complements, so a factor has the largest effect if it is equated last. Put differently, reducing friending bias leads to more high-SES friends when exposure is higher (and vice versa).

Decompositions Across Areas

We use a similar approach to analyze why economic connectedness among low-SES people varies geographically. We begin by calculating, for every ZIP code a, average friending shares by setting, average friending bias ($\text{Friending Bias}_{H,L,a}$) and average high-SES exposure ($\text{Exposure}_{H,L,a}$) of low-SES individuals living in that area (see Supplementary Information B.5 for formal definitions). We then consider representative agents with friending shares across settings and setting-specific levels of exposure and friending bias that match the mean values of these parameters for low-SES individuals living in ZIP codes in the bottom vs. top quintiles of the ZIP code-level EC distribution. We then sequentially set each of the parameters for the representative bottom-ZIP-quintile agent to match the values for the representative top-ZIP-quintile agent (see Supplementary Information B.5 for details).

VI.E Exposure, Bias, and Upward Income Mobility

In Table 2, we analyze the relationship between upward income mobility and economic connectedness across counties and ZIP codes, comparing the effects of exposure vs. friending bias. Starting from the area-level mean values for exposure and friending bias among low-SES individuals (see the Decompositions Across Areas section in Section VI.D), we first create a recomposed measure of EC based on the product of the average values of exposure and one minus friending bias in each area a:

$$\text{EC}^{rec}_{H,L,a} = \text{Exposure}_{H,L,a} \times (1 - \text{Friending Bias}_{H,L,a}). \tag{6}$$

Note that $\text{EC}^{rec}_{H,L,a}$ differs from the measures of area-level EC analyzed in Chetty et al. (2022) because (1) the measure here is only based on the subset friendships that can be assigned to groups and (2) the product of the area-level averages does not take into account the covariances between friend shares, exposure, and bias at the individual level. Nevertheless, the two measures of area-level EC have a population-weighted correlation of above 0.95 across both counties and ZIP codes.

Because economic connectedness is proportional to the *product* of exposure and one minus friending bias, we use a log transformation of equation (6) to obtain an additive specification:

$$\ln(EC^{rec}_{H,L,a}) = \ln(\text{Exposure}_{H,L,a}) + \ln(1 - \text{Friending Bias}_{H,L,a}). \tag{7}$$

We then regress the log of upward mobility at the county and ZIP code level on these log-transformed measures of economic connectedness or exposure and friending bias in columns 1-6 of Table 2, weighting by the below-median-SES population. In column 7 of Table 2, we define the dependent variable as the log of the causal effect of a county on upward mobility, which we calculate as the average level of upward mobility overall in the U.S. plus twenty times Chetty and Hendren's (Chetty and Hendren 2018) estimate of the annual causal exposure effect of growing up in that county.

VI.F High School Estimates

For both high-school- and college-level estimates of friending bias and exposure using own SES, we focus on the 1986-1996 birth cohorts (measuring SES in 2022, between the ages of 26-36). For estimates based on parental SES, we focus on individuals in the 1990-2000 birth cohorts. We focus on more recent birth cohorts for parental SES in order to maximize the fraction of individuals we can link to their parents and to measure parental SES (in 2022) before many parents begin to retire. For the 1990-2000 cohorts, we are able to link 46% of individuals assigned to high schools to parents with a non-missing SES rank. We pool data over several cohorts to obtain more precise estimates. School-level estimates of economic connectedness are generally stable over time; for example, across schools, EC in the 1978-1982 birth cohorts has a correlation of 0.87 with EC in the 1993-1997 cohorts (Supplementary Figure 5).

To estimate the reliabilities of exposure and bias for high schools, we first randomly split the population of each high school into two subpopulations, and compute exposure and bias on the subgraphs formed by these two populations. We then take a weighted correlation of these exposure or bias statistics across the split samples, weighting by the number of low-SES students in the school. To adjust for the fact that the estimates are naturally more noisy when estimated on only half of the sample, as opposed to the full sample that we actually use to construct our baseline estimates, we divide the raw split-sample correlation coefficient by the ratio of the (weighted) full-sample variance of EC across schools to the (weighted) split-sample variance of EC across schools.

High-parental-SES individuals make about 22% more friends in high school than low-parental-SES individuals do (Chetty et al. 2022). Hence, applying equation (5), we would expect friending bias of -0.10 in a school where 50% of students have high parental SES and friendships exhibit no homophily, but high-parental-SES students continue to make 22% more friends than low-parental-SES students.

VI.G Total Contribution to Connectedness

We define each school g's *total contribution to economic connectedness (TCEC)* as:

$$\text{TCEC}_g = (1 - \text{Exposure}_g) \times \text{Exposure}_g \times (1 - \text{Friending Bias}_g),$$

where Exposure_g and Friending Bias_g are the average high-SES exposure and friending bias of low-SES students in school g. In this equation,$\text{Exposure}_g \times (1 - \text{Friending Bias}_g) \approx EC_g$, where EC_g is the average economic connectedness of low-SES students in the school. Note that this equality only holds approximately because of the potential covariance between exposure and bias across cohorts within a school. For similar reasons, $(1 - \text{Exposure}_g) \times EC_g$ is only approximately equal to the total number of cross-SES links formed per student. Abstracting from any such covariance between exposure and friending bias, $TCEC_g$ measures a school's overall contribution to economic connectedness per student.

VI.H Cross-Cohort Estimates

Under the identification assumption that fluctuations in peer group composition across cohorts are orthogonal to other unobservable determinants of students' friending choices, fluctuations in the share of high-SES peers across cohorts within a school can be used to identify the causal effect of exposure on economic connectedness. Prior work in the peer effects literature has found support for this identification assumption using a variety of balance and placebo tests (Sacerdote 2011).

To implement the cross-cohort research design, we begin by assigning each person born between 1990 and 2000 to a high school cohort based on their high school and birth date. We use parental SES for this analysis, which—unlike children's own future SES in adulthood—is exogenous to one's high school peer group. Because the design relies on small-sample variation in exposure, we focus on connections between individuals with parents in the lowest and highest SES quintiles (rather than below vs. above median SES) to increase variation.

In analogy to a hypothetical experiment that randomly increases the number of high-SES students in a given cohort c, let $\Delta\text{Exposure}_{sc}$ denote the share of high-SES (top-parental-SES-quintile) peers in school s in cohort c minus the mean share of high-SES peers in school s in all other cohorts excluding c (divided by the top-quintile population share, 20%). Similarly, let ΔEC_{sc} denote the difference in economic connectedness between cohort c and the mean economic connectedness of all other cohorts in the same school. Here, we measure economic connectedness within cohorts—i.e., the share of high-SES friends that low-SES students make within their cohorts in their high schools. Figure 9a presents a binned scatter plot of ΔEC_{sc} vs. $\Delta\text{Exposure}_{sc}$, pooling all schools.

To construct Figure 9b, we first divide school-cohort cells into deciles based on the average level

of friending bias in all other cohorts in the same school *leaving out* the focal cohort c. We use this leave-out approach to mirror the decision problem of a principal who uses data on existing high school cohorts to estimate school-level bias, and then uses that estimate to predict the effects of future changes in exposure on economic connectedness. We estimate a regression analogous to that shown in Figure 9a separately for cohorts in each decile of the friending bias distribution. Figure 9b plots the estimated regression coefficients in each decile against the level of leave-out friending bias (based on all other cohorts excluding the focal cohort) in that decile.

We use all other cohorts (including future cohorts) to maximize precision when estimating friending bias in Figure 9, but obtain similar results when we only use prior cohorts to calculate school level friending bias for a given cohort c. We also obtain similar estimates when we use first-differences instead of fixed effects (i.e., comparing changes in EC and exposure across adjacent cohorts instead of demeaning with respect to all cohorts in the school) and when demeaning EC and exposure in each cohort with respect to the two neighboring cohorts rather than all cohorts (Supplementary Figure 6).

VI.I Regression Discontinuity Estimates

The regression discontinuity design induces quasi-random assignment across adjacent cohorts, and thus addresses potential biases that could arise in the cross-cohort design from correlated trends in exposure and EC. The identification assumption underlying the RD design is that other determinants of friending behavior do not jump discretely across cohorts in a manner that is correlated with differences in the share of high-SES students across cohorts. We assess the validity of this assumption in Supplementary Information B.6, where we show that observable characteristics do not jump discretely at cohort cutoffs with changes in exposure.

As in the cross-cohort design, we focus on within-cohort friendships between bottom-parental-SES-quintile and top-parental-SES-quintile students. To implement the design, we begin by focusing on pairs of adjacent cohorts where the magnitude of the jump in top-parental-SES-quintile (high-SES) exposure $|\Delta E_{sc}| = \text{Exposure}_{sc} - \text{Exposure}_{s,c-1}$ lies in the top quartile of the distribution of $|\Delta E_{sc}|$ across cohort pairs. On average, high-SES exposure jumps by approximately 0.40 units at the entry date cutoff, pooling all cohort pairs in the top quartile of $|\Delta E_{sc}|$.

In Figure 10a, we examine how these jumps in exposure to high-parental-SES peers affect individual economic connectedness in schools with low (bottom quartile) vs. high (top quartile) friending bias. As above, we use leave-out estimates of friending bias for each cohort calculated as the average friending bias in all other cohorts of the same school, excluding the two focal cohorts used in the RD analysis. Each series in the figure plots within-cohort EC—the share of high-parental-SES friends that low-

parental-SES students have within their high school cohort divided by 0.2, the high-SES population share—vs. their date of birth, subtracting the prior cohort mean to isolate changes across cohorts.

For each friending bias quartile, we estimate the magnitude of the jump in EC at the entry cutoff by regressing cohort-specific EC on date of birth, an indicator for being above the entry date cutoff, and the interaction of date of birth with the indicator for being above the entry date cutoff (see Supplementary Information B.6 for the estimating equation). We use a bandwidth of 200 days around either side of the cutoff in this regression; we show the robustness of our estimates to other bandwidths in Supplementary Information B.6.

Figure 10b collects the RD estimates obtained from analogous regression specifications for all four quartiles of the distribution of $|\Delta E_{sc}|$. It then plots those estimates vs. the mean change in $|\Delta E_{sc}|$ in each quartile, separately for schools in the top vs. bottom quartile of friending bias.

VI.J Privacy and Ethics

The project focuses on drawing high-level insights about communities and groups of people, rather than individuals. We used a server-side analysis script that was designed to automatically process the raw data, strip the data of personal identifiers, and generate aggregated results, which we analyzed to produce the conclusions in this paper. The script then promptly deleted the raw data generated for this project. While we used various publicly available sources of aggregate statistics to supplement our analysis, we do not link any external individual-level information to the Facebook data. All inferences made as part of this research were created and used solely for the purpose of this research and were not used by Meta for any other purpose.

A publicly available dataset, which only includes aggregate statistics on social capital, is available on www.socialcapital.org. We use methods from the differential privacy literature to add noise to these aggregate statistics to protect privacy while maintaining a high level of statistical reliability; see www.socialcapital.org for further details on these procedures. The project was approved under Harvard University IRB 17-1692.

Supplementary Information

A Supplementary Information on Data

In this section, we describe the external (non-Facebook) data we use in our analysis. Note that we do not link any external individual-level information to the Facebook data.

A.1 Upward Mobility

Data on economic mobility by Census tract and county are obtained from the publicly available Opportunity Atlas (Chetty et al. 2018). We define upward income mobility in each area as the average income percentile in adulthood of a child born to parents at the 25th percentile of the income distribution. We aggregate the Census tract data on upward mobility to the ZIP code (ZCTA) level using the number of children with below-median parental income as weights.

A.2 High School Characteristics

National Center for Education Statistics (NCES). Data on high schools and their characteristics are publicly available from the National Center for Education Statistics (NCES), the primary federal entity for collecting data from U.S. high schools. We obtain measures of the total number of students in grades 9 through 12, the percent of students eligible for free or reduced lunch (for public schools), and racial shares (Black, white, Asian, Hispanic, Native American) from the 2017-2018 Common Core of Data and Private School Universe Survey. We also use the surveys conducted in 1997-1998 and 2007-2008 to obtain lists of schools that have since closed, which we use to match individuals to schools used in our analyses (note, however, that we do not publicly release statistics on social capital for schools that have closed).

Civil Rights Data Collection (CRDC). We supplement the NCES data with data from 2015–2016 Civil Rights Data Collection (CRDC), a biennial survey of public schools on various civil rights indicators related to access to education. We obtain measures of the share of students in AP courses and the share of students enrolled in gifted or talented programs from the CRDC.

A.3 College Characteristics

Integrated Postsecondary Education Data System (IPEDS). Data on colleges and their characteristics, including racial shares (Black, white, Asian or Pacific Islander, Hispanic), are obtained from the 2013 Integrated Postsecondary Education Data System.

College Mobility Report Cards (Tax Data). We supplement the IPEDS data with data on the incomes of college students' parents, which are obtained from variables that were publicly released by Chetty et al. (2020). We also use the college tier variable from this dataset, which was constructed using IPEDS data.

B Supplementary Methods

This section provides details on six aspects of our methods: (1) identifying the groups in which individuals participate; (2) allocating friendships to these groups; (3) adjusting for under-reporting of group memberships; (4) the treatment of zero friend shares and zero exposure in decompositions, (5) formal definitions of various statistics shown in figures and tables; and (6) details on the regression discontinuity design used to identify the causal effects of changes in exposure on EC. Details on other aspects of our methods—such as estimation of socioeconomic status and computation of standard errors—are available in Supplementary Information B of Chetty et al. (2022).

B.1 Assigning Facebook Users to Groups

We begin by identifying the groups to which individuals in our analysis sample belong, focusing on the following six settings where friendships are formed: high schools, colleges, recreational groups, religious groups, workplaces, and neighborhoods (defined as ZIP codes). We describe how we assign individuals to specific groups in each of these settings in turn.

High schools. We assign individuals to high schools based on self-reported high schools, self-reported hometowns, and information from their social networks. We begin by matching self-reported high school names to the National Center for Education Statistics' (NCES) comprehensive surveys of U.S. public and private schools. We drop virtual schools, schools located in the five U.S. territories or on military bases abroad, and schools with fewer than 50 students. For individuals who self-report a common high school name (e.g., "Central High School") we only include that self-report if the individual also reports a hometown that matches the school's location. For the 3.3% of individuals with multiple self-reported high schools, we assign the school at which the individual has the greatest number of friends whose ages are within three years of the individual's own age. We exclude self-reported schools where individuals have fewer than 10 friends with ages within three years of their own (i,e., we do not include them when calculating EC, exposure, or bias for that school); however, we include these individuals in the potential set of friends for other individuals assigned to that school.

For people without a validated self-reported high school, we use their friendship network to impute their high school. For this imputation, we only consider friends who have a valid self-reported high school and who are within three years of the individual's age. We then calculate the ratio of an individual's friends in the high school where they have the most friends relative to the schools where they have the next most friends, and assign the user to the first high school if this ratio exceeds two (we further require that the individual has at least five friends in the first high school). We evaluate the accuracy of this imputation approach using the sample of users with validated self-reports. For users with a valid self-reported high school, the network-imputed high school matches the self-reported high school 97.4% of the time. Using this algorithm, we observe high schools for 74.9% of individuals in our analysis sample; 53.8% are assigned via self-reports and 21.1% via imputation based on their friendship network.

For the quasi-experimental analysis in the "Effects of Integration on Connectedness" section, we assign students to high school cohorts by collecting data on school entry cutoffs by year from Elder and Lubotsky (2009) and Bush and Zinth (2011) (in all other analyses, we define individuals' high school cohorts simply based on calendar years of birth for simplicity). See Supplementary Table 2 for a list of the entry date cutoffs we use in our analysis. The average high school cohort has 115 users assigned to it (Supplementary Table 3). To ensure that our quasi-experimental estimates are not biased by imputing high schools based on friendship networks, we verify that the results in Figure 10 remain similar when restricting to individuals for whom we observe high schools based solely on self-reports.

Colleges. To assign people to colleges, we begin by matching individuals' self-reported colleges to the IPEDS directory. We drop online colleges as well as those that do not appear in the Carnegie Classification. For the 17.8% of individuals with multiple self-reported colleges conditional on being assigned a college, we use the one with the maximum number of friends (restricting attention friends within three years of the individual's age who have a valid self-reported college). We exclude self-reported colleges where individuals have fewer than 10 friends with ages within three years of their own (i,e., we do not include them when calculating EC, exposure, or bias for that college); however, we include these individuals in the potential set of friends for other individuals assigned to that college. We observe a college for 42.9% of individuals in our analysis sample, and the average college has 530 users per cohort.

Recreational groups. To analyze friendships formed in recreational groups, we use algorithms that classify Facebook groups by topic based on their titles. Since our goal is to capture recreational

activities that can facilitate real-life friendships, we restrict attention to groups classified as sports, fitness, performing arts, crafts, and literature. We also exclude any groups related to the buying and selling of items. Among these groups, we impose two further restrictions to increase the likelihood that members of the group have met in person (i.e., not just virtually). First, we only consider groups with between 10 and 3,000 members in our primary analysis sample. Second, we require that at least 80% of group members must reside in a single commuting zone. We exclude members who do not live in this modal commuting zone. For the 45.9% of users belonging to multiple recreational groups that satisfy all of these restrictions, we select the recreational group with the maximum number of friends.[7] We assign 29.8% of individuals in our analysis sample a recreational group using this approach, each of which has 17 users on average assigned to it.

Faith-based (Religious) Groups. We use regular expressions to identify Facebook pages that are faith-based (excluding pages containing education, conference, event, media, or music-related phrases), and restrict attention to pages with an admin-reported U.S. address. We also identify faith-based Facebook groups, which are classified based on the title of the group and other group characteristics. When identifying faith-based pages or groups, we only consider pages with between 20 and 2,000 likes and groups with between 20 and 2,000 members.

We assign individuals to pages based on page likes and to groups based on the individuals joining the particular group. We only assign individuals to pages or groups that are located in the individual's own commuting zone. In the case of assignment to multiple pages or groups, we prioritize those located in the individual's county, and within that subset select the page or group with the highest number of friends. We identify a faith-based page or group for 17.9% of users using this approach. On average, each faith-based page or group has 39 individuals in our analysis sample.

For convenience, we refer to faith-based groups identified using this approach as "religious groups" throughout the paper. Note that our classification does not infer an individual's religion: it is focused on whether friendships were formed in a faith-based group, rather than whether an individual belonged to or identified with a particular religion.

Workplaces. We use self-reported data on individuals' Facebook profiles to assign them to workplaces. We apply three restrictions to increase data quality. First, we remove self-reported employers that are clearly fictional or contrived (e.g., "The Krusty Krab", "None of your business", and "Businessperson"). Next, we remove employment that does not involve external interactions such as self-employment, stay-at-home parenting, and blogging. Finally, we require the linked employer page to contain at least 10 other employees between the ages of 22 and 65 from the same county as the user (but not the same county as the employer page, since regional branches may not be listed separately). For the 9.5% of users with multiple valid employers (among those with at least one valid employer), we select the employer with the highest number of friends.

We observe an employer for 20.9% of individuals in our analysis sample. Each employer has on average 45 users assigned to them. As expected given the skewed distribution of the size of firms, some employers are matched to far more people than others; the distribution exhibits substantial dispersion, with a standard deviation of 881.

Neighborhoods. We use ZIP code tabulation areas (ZIP codes) to represent neighborhoods. Because Facebook users only have one current ZIP code, they are mechanically matched to a single neighborhood. If an individual has fewer than 10 friends in their assigned neighborhood, we define their neighborhood as missing (i,e., we do not include them when calculating EC, exposure, or bias for that neighborhood); however, we include these individuals in the potential set of friends for other individuals assigned to that neighborhood.

[7] As a robustness check to the method of selecting only the top group by friendship links, we also calculate metrics of friendship behavior using two alternative approaches: taking the union of all groups and random assignment of groups. For recreational groups, the individual-level EC correlations between just taking the top group vs. the union of all groups and random assignment of groups are 0.98 and 0.96. The corresponding individual-level group bias correlations are 0.88 and 0.85. We perform the same exercise for other settings and find similarly high correlations.

Note that when we measure economic connectedness for neighborhoods in this paper, we refer to the share of high-SES friends among friends who live *within* an individual's own ZIP code. This differs from the neighborhood-level economic connectedness measures analyzed in our first paper (Chetty et al. 2022), which are based on *all* the friends of people living in a given neighborhood, irrespective of whether those friends live in the same neighborhood.

Supplementary Table 3 shows statistics on the share of individuals we are able to assign to a group in each of these settings, both overall and by SES. Across all settings (except neighborhoods), we assign high-SES individuals to groups at a higher rate than low-SES groups. These differences are largely due to differences in participation rates by SES; for instance, high-SES people are more likely to have attended college and be working, and hence it is natural that they are more likely to be assigned to colleges or employers. Consistent with this view, our statistics on the share of high-SES individuals by group are highly correlated with external, publicly available data (Table 3). Nevertheless, some of the differences in rates of group membership by SES may be driven by differences in reporting rates. To assess whether such reporting differences might bias our results, we develop an algorithm to adjust for differential under-reporting by SES using external data on group memberships. we describe this algorithm in Section B.3 below, and show that our results remain similar with this correction.

B.2 Assigning Friendships to Settings

After assigning individuals to groups, we allocate their friendships to these groups. A friendship is assigned to a particular group (e.g., a high school or a recreational group) when both individuals in the friendship are members of the group.

When friends appear in multiple settings—for example, when two friends share a common recreational group and a common employer—one could either attempt to assign the friendship to the setting that sparked the friendship or assign the friendship to all settings in which it appears. We adopt the latter agnostic approach. For example, if a user and their friend share both a high school and a recreational group, then that friendship link is counted in both settings. While this approach overweights friendships that appear in multiple settings, it avoids the need to adjudicate between settings and may appropriately place more weight on more important friends with whom one has contact in multiple settings. Using this procedure, we assign 30% of friendships to at least one of the six settings.

In Supplementary Table 3b, we report the percentage of friends from each setting that overlaps with another setting. For instance, 27.6% of college friends are also high school friends, and 36.1% of recreational group friends are neighborhood friends (i.e., they live in the same ZIP code).

B.3 Adjusting for Underreporting

The procedure described above to assign friendships to groups likely underestimates the raw number of friends made in each setting because we rely on user self-reports and conservative imputation procedures. We evaluate the sensitivity of our results to such under-reporting by implementing adjustments for the number of friends made in each setting in two steps.

First, we estimate reporting rates for each setting by dividing the share of individuals assigned to a setting in the Facebook data by the "true" share of individuals in the population who participate in that setting based on external data sources (e.g., the share of people who are members of religious groups). For high school attendance, college attendance, and employment rates, we use data from the 2014–2018 American Community Survey 5-year estimates (U.S. Census Bureau 2017) at the state $\times$ SES level. We use graduation rather than attendance rates to benchmark high school and college participation because Facebook users might be more likely to report their high school/college if they graduated from it. For religious group participation, we use data from the Pew Research Center (Pew Research Center 2014) at the state level. For neighborhoods, there is no under-reporting because we observe ZIP codes for everyone. We are unable to obtain reliable

estimates of true participation rates for recreational groups and therefore do not make adjustments for those groups. We measure participation rates at the finest available geographic/SES level for each setting. For example, we calculate reporting rates at the state × SES (above/below median) level for colleges, but we are only able to calculate a state-level reporting rate for religious groups as we do not have data on religious group attendance by SES and state.

Second, we use these reporting rates to correct the average share of friends at the setting × state × SES level. For each state t and SES type e, we define an adjusted share of friends in each setting s as:

$$\text{Adjusted Friend Share}_{t,e,s} \equiv \frac{\text{Raw Friend Share}_{t,e,s}}{r_{t,e,s} E[r_{t,e(j),s} | j \in N_{t,e,s}]}, \tag{8}$$

where $r_{t,e,s}$ is the reporting rate estimated for state t, SES type e, and setting s and $E[r_{t,e(j),s}|j \in N_{t,e,s}]$ is the average reporting rate in that setting for the set of friends $N_{t,e,s}$ made in that setting by people of SES type e (noting that within a setting, high and low-SES friends may have differing reporting rates). Intuitively, we inflate raw friend shares by the product of these two reporting rates to account for the fact that a friendship link is missed if either the individual or the friend does not report their group membership. Finally, we normalize the adjusted shares of friends across the six settings to sum to 1 within each state × SES type.

In Supplementary Figure 4, we use these adjusted friend shares instead of the raw friend shares as weights in the decompositions of differences in economic connectedness. The results are similar to those obtained with the raw shares used in our baseline analysis, suggesting that underreporting of friendships does not significantly bias our estimates.

B.4 Treatment of Zero Friend Shares in Decompositions

In this section, we describe how we define the share of high-SES friends $f_{H,i,g}$ in groups where an individual i makes no friends for the decompositions discussed in Section VI.D. Formally, $f_{H,i,g}$ is undefined in groups that i does not belong to. However, since $\phi_{i,g} = 0$ for any such group, our calculations are unaffected by how $f_{H,i,g}$ is defined for those groups; we therefore set $f_{H,i,g} = 0$ for groups g of which i is not a member.

We also exclude groups having $w_{H,g} = 0$ from G because friending bias $(1 - \frac{f_{H,i,g}}{w_{H,g}})$ is not well-defined for such groups. Only 0.01% of group memberships have $w_{H,g} = 0$ when "high-SES" is defined as having above-median SES; thus, our results would be very similar if we were to use a different convention in such cases.

B.5 Definitions of Statistics Reported in Tables and Figures

In this section, we provide formal definitions of the statistics reported in Figures 1–5 and 9, which report and decompose measures of economic connectedness, exposure, and friending bias at different levels of aggregation. Let i index individuals in our sample and g index groups. Let $G(i)$ denote the set of groups of which i is a member and in which i makes at least one friend.

Figure 1: Figure 1 plots friending shares by setting for individuals with different SES. The settings are $s \in$\{college, employer, high school, neighborhood, recreational group, religious group\}. Individuals can only be part of one group in a given setting (e.g., individuals can only be assigned one high school) and the share of friends made in group g is $\phi_{i,g} = 0$ for all groups g of which i is not a member. We compute average friending rates $\phi_{v,s}$ in each setting s separately for individuals in different SES ventiles $v \in \{1, 2, \ldots, 20\}$. Let $v(i)$ denote i's SES ventile and $s(g)$ denote the setting of a group g. Then, setting-level friending shares by SES ventile are:

$$\phi_{v,s} = \frac{1}{N_v} \sum_{v(i)=v} \sum_{s(g)=s} \phi_{i,g}, \tag{9}$$

where N_v denotes the number of individuals who are in SES ventile v. Figure 1 plots these friending shares $\phi_{v,s}$ normalized by the overall mean friend shares in each setting. That is, for each SES ventile v and setting s, we plot $\phi_{v,s}/(\sum_{v=1}^{20} \phi_{v,s}/20)$.

Figure 2: Figure 2 plots mean EC, exposure, and friending bias by setting, separately for individuals with below-median-SES ($SES = L$) and above-median-SES ($SES = H$). In what follows, note that we define the share of high-SES friends $f_{H,i,g} = 0$ for all groups g of which i is not a member (see Supplementary Information B.4 above). Let $N_{SES,s}$ denote the number of individuals who have a certain SES and make at least one friend in setting s, e.g., $N_{L,college}$ refers to the number of low-SES individuals in our analysis sample who make at least one friend in college.

Panel A: We compute average economic connectedness to high-SES individuals, $\text{EC}_{H,SES,s}$, separately for $SES \in \{L, H\}$ types in each setting s. These values are computed by averaging individuals' shares of high-SES friends, corresponding to $f_{H,i,g}$ from equation (2) in Methods, in the groups that belong to setting s, i.e. $s(g) = s$, over all individuals of type SES who are assigned to a group in setting s. This expression is then normalized by the share of high-SES individuals in the national population, $w_H = 0.5$:

$$\text{EC}_{H,SES,s} = \frac{1}{N_{SES,s} w_H} \sum_{i \in SES} \sum_{s(g)=s} f_{H,i,g}. \tag{10}$$

The bar chart plots $\text{EC}_{H,L,s}$ and $\text{EC}_{H,H,s}$ separately for each setting.

Panel B: We compute average exposure to high-SES individuals, $\text{Exposure}_{H,SES,s}$, separately for $SES \in \{L, H\}$ types in each setting s. Note that because we require individuals to attend high school or college within three cohorts of each other in order to be deemed high school or college friends, we measure individuals' exposure to high-SES peers in their high schools and colleges to be the share of high-SES individuals in the cohorts that lie within three years of the individual's own birth cohort (defining cohorts based on calendar years). In these cases, exposure therefore depends on the particular cohort that an individual is in. For example, the exposure in college of a student in cohort c, which we denote by $\text{Exposure}^c_{H,\text{college}}$, is twice the share of high-SES students in cohorts $\{c-3, c-2, \ldots, c+3\}$. Let $c(i)$ denote individual i's cohort and $w^{c(i)}_{H,g}$ denote the share of high-SES individuals in the relevant set of three cohorts around c.[8] In the other settings (neighborhoods, workplaces, recreational and religious groups), we do not impose cohort restrictions to assign friendships, and hence every group member has the same exposure and the cohort superscript can be ignored.

We define $\text{Exposure}_{H,SES,s}$ as the mean share of high-SES peers in the group g to which individual i belongs in setting ($s(g) = s$), averaging over all individuals of type SES who have at least one friend in setting s. We then normalize this expression by the share of high-SES individuals in the national population, $w_H = 0.5$:

$$\text{Exposure}_{H,SES,s} = \frac{1}{N_{SES,s} w_H} \sum_{i \in SES} \sum_{\substack{g \in G(i) \\ s(g)=s}} w^{c(i)}_{H,g}. \tag{11}$$

The bar chart plots $\text{Exposure}_{H,L,s}$ and $\text{Exposure}_{H,H,s}$ separately for each setting.

Panel C: We compute average friending bias with respect to high-SES individuals, $\text{Friending Bias}_{H,SES,s}$, separately for $SES \in \{L, H\}$ types in each setting s. At the individual level, friending bias is defined as the ratio of an individual's share of high-SES friends to their share of high-SES peers $\frac{f_{H,i,g}}{w^{c(i)}_{H,g}}$

[8] We also measure economic connectedness and friending bias only based on friends in an individual's cohort and the three surrounding cohorts. However, since the high-SES friend share $f_{H,i,g}$ is already defined at the individual level rather than the group level, we can simplify notation and do not include an additional superscript c for those expressions.

in an individual's group that belongs to setting s ($s(g) = s$). We then compute mean friending bias by SES type and setting by averaging individual friending bias over all individuals of type SES who have at least one friend in setting s:

$$\text{Friending Bias}_{H,SES,s} = 1 - \frac{1}{N_{SES,s}} \sum_{i \in SES} \sum_{\substack{g \in G(i) \\ s(g)=s}} \frac{f_{H,i,g}}{w_{H,g}^{c(i)}}. \tag{12}$$

The bar chart plots $\text{Friending Bias}_{H,L,s}$ and $\text{Friending Bias}_{H,H,s}$ separately for each setting.

Panel D: As in Panel C, we first compute average friending bias with respect to high-SES individuals, $\text{Friending Bias}_{H,L,s}$, for low-SES individuals in each setting s, restricting the sample to members of a religious group. Let R be the set of individuals i who are members of religious groups (i.e., any i such that $\exists g \in G(i)$ with $s(g) =$ religious group). Let $N_{L,s,R}$ denote the number of low-SES religious group members who are also part of a setting s and make at least one friend in setting s. The values in the bars are given by computing mean friending bias in each setting as in Panel C using the subset of low-SES individuals who are part of a religious group ($i \in L \cap R$), and then subtracting the mean friending bias in the religious group setting for those individuals. Formally, we plot the following for all non-religious group settings s, using "rel" as an abbreviation for "religious group":

$$\begin{aligned}
&\text{Friending Bias}_{H,L,s,R} - \text{Friending Bias}_{H,L,\text{rel}} \\
&= \left(1 - \frac{1}{N_{L,s,R}} \sum_{i \in L \cap R} \sum_{\substack{g \in G(i) \\ s(g)=s}} \frac{f_{H,i,g}}{w_{H,g}^{c(i)}}\right) - \left(1 - \frac{1}{N_{L,\text{rel}}} \sum_{i \in L \cap R} \sum_{\substack{g \in G(i) \\ s(g)=\text{rel}}} \frac{f_{H,i,g}}{w_{H,g}}\right) \\
&= \frac{1}{N_{L,\text{rel}}} \sum_{i \in L \cap R} \sum_{\substack{g \in G(i) \\ s(g)=\text{rel}}} \frac{f_{H,i,g}}{w_{H,g}} - \frac{1}{N_{L,s,R}} \sum_{i \in L \cap R} \sum_{\substack{g \in G(i) \\ s(g)=s}} \frac{f_{H,i,g}}{w_{H,g}^{c(i)}}
\end{aligned} \tag{13}$$

Figure 3A: To construct Figure 3A, we begin by defining a representative-agent notion of economic connectedness based on the product of the setting-specific average friending rates, average exposure, and average bias for people in that SES group:

$$EC_{H,SES} \equiv \sum_s \phi_{SES,s} \times \text{Exposure}_{H,SES,s} \times (1 - \text{Friending Bias}_{H,SES,s}), \tag{14}$$

where the variables $\text{Exposure}_{H,SES,s}$ and $\text{Friending Bias}_{H,SES,s}$ are defined as in equations (11) and (12) above, and $\phi_{SES,s} = (1/N_{SES}) \sum_{i \in SES} \sum_{s(g)=s} \phi_{i,g}$. This equation is a representative-agent analog of equation (2) in Methods, in which $EC_{H,SES}$ is the product of setting-level means of friending shares, exposure, and bias, which can differ from an average of individual-level EC. $EC_{H,SES}$ may also differ from the across-setting mean of $EC_{H,SES,s}$ as defined above. The reason is that multiplying setting-level averages ignores any covariance between friend shares, exposure, and friending bias at the individual level.[9]

We use equation (14) to decompose why $EC_{L,SES}$ differs from $EC_{H,SES}$.

Bar 1: This bar shows $EC_{H,L}$ as defined in equation (14).

[9] As a reference, the top bar in Figure 3A shows that the EC of low-SES individuals is 0.83 on average, while the bottom bar shows that the corresponding value is 1.53 for high-SES individuals—corresponding to a gap in EC of 0.7. In addition to the differences driven by the covariance terms just described, both numbers also differ slightly from the EC gap between low- and high-SES individuals of $1.4 - 0.78 = 0.63$ reported in Chetty et al. (2022), since that number is based on all friends, while the numbers in this paper are only based on friendships that we can allocate to a group where they were likely formed.

Bar 2: In this bar, we set the friending shares of low-SES users to be equal to those of high-SES individuals, i.e., $\forall s$, set $\phi_{L,s} = \phi_{H,s}$. As such, this bar plots $\sum_s \phi_{H,s} \times \text{Exposure}_{H,L,s} \times (1 - \text{Friending Bias}_{H,L,s})$.

Bar 3: In this bar, we keep the equated friending shares and also equate the setting-level exposure of low-SES individuals to that of high-SES individuals, i.e., $\forall s$, set $\text{Exposure}_{H,L,s} = \text{Exposure}_{H,H,s}$. This bar plots $\sum_s \phi_{H,s} \times \text{Exposure}_{H,H,s} \times (1 - \text{Friending Bias}_{H,L,s})$.

Bar 4: In this bar, instead of equating exposure, we equate bias, i.e., $\forall s$, set $\text{Friending Bias}_{H,L,s} = \text{Friending Bias}_{H,H,s}$, and plot $\sum_s \phi_{H,s} \times \text{Exposure}_{H,L,s} \times (1 - \text{Friending Bias}_{H,H,s})$.

Bar 5: Lastly, in the bottom bar, we report $EC_{H,H}$ as defined in equation (14), which is equivalent to equating friending shares, exposure, and bias simultaneously, i.e., $\sum_s \phi_{H,s} \times \text{Exposure}_{H,H,s} \times (1 - \text{Friending Bias}_{H,H,s}) = EC_{H,H}$.

Figure 3B: In Figure 3B, we conduct a similar counterfactual exercise across ZIP codes. We decompose the economic connectedness in ZIP codes with the highest vs. the lowest 20% of average economic connectedness among low-SES individuals to understand what drives differences in EC among low-SES individuals across ZIP codes. To this end, we define representative-agent measures of mean friending shares, exposure, and bias for low-SES individuals in different ZIP-code-EC-quintiles and settings. Let $z(i)$ denote i's ZIP code.
We divide ZIP codes z into quintiles $q \in \{1, 2, \ldots, 5\}$ in terms of their ZIP-code-level economic connectedness, $EC_{H,L,z} = \frac{1}{N_{L,z} w_H} \sum_{\substack{i \in L \\ z(i)=z}} \sum_{g \in G(i)} \phi_{i,g} \times f_{H,i,g}$, weighted by the number of low-SES individuals in each ZIP code. Let $q(z)$ denote the EC quintile of ZIP code z. Let $N_{L,q}$ denote the number of low-SES individuals living in ZIP codes that are in EC quintile q and $N_{L,s,q}$ the number of low-SES individuals in those ZIP codes who additionally make at least one friend in setting s. We then define the following setting $\times$ quintile measures of friend shares, exposure, and friending bias among low-SES individuals:

$$\phi_{L,s,q} = \frac{1}{N_{L,q}} \sum_{\substack{i \in L \\ (q \circ z)(i)=q}} \sum_{s(g)=s} \phi_{i,g} \tag{15}$$

$$\text{Exposure}_{H,L,s,q} = \frac{1}{N_{L,s,q} w_H} \sum_{\substack{i \in L \\ (q \circ z)(i)=q}} \sum_{\substack{g \in G(i) \\ s(g)=s}} w_{H,g}^{c(i)}, \tag{16}$$

$$\text{Friending Bias}_{H,L,s,q} = 1 - \frac{1}{N_{L,s,q}} \sum_{\substack{i \in L \\ (q \circ z)(i)=q}} \sum_{\substack{g \in G(i) \\ s(g)=s}} \frac{f_{H,i,g}}{w_{H,g}^{c(i)}}. \tag{17}$$

We then decompose the following equation:

$$EC_{H,L,q} \equiv \sum_s \phi_{L,s,q} \times \text{Exposure}_{H,L,s,q} \times (1 - \text{Friending Bias}_{H,L,s,q}). \tag{18}$$

Similar to the previous decomposition, note that the representative-agent measure $EC_{H,L,q}$ here may differ from the average of individual-level $IEC_{H,i}$ among low-SES individuals across ZIP codes in the bottom quintile of average economic connectedness, because of individual-level covariances between friend shares, exposure, and friending bias.
We use equation (18) to decompose why $EC_{H,L,1}$ differs from $EC_{H,L,5}$.

Bar 1: This bar shows $EC_{H,L,1}$ as defined in equation (18).

Bar 2: In this bar, we set the friending shares of low-SES individuals in low-EC ZIP codes to be those of low-SES individuals in high-EC ZIP codes, i.e., $\forall s$, set $\phi_{L,s,1} = \phi_{L,s,5}$. As such, this bar shows $\sum_s \phi_{L,s,5} \times \text{Exposure}_{H,L,s,1} \times (1 - \text{Friending Bias}_{H,L,s,1})$.

Bar 3: In this bar, we keep the equated friending shares and also equate the setting-level exposures of low-SES individuals in low-EC ZIP codes to those of low-SES individuals in high-EC ZIP codes, i.e., $\forall s$, set $\text{Exposure}_{H,L,s,1} = \text{Exposure}_{H,L,s,5}$. This bar plots $\sum_s \phi_{L,s,5} \times \text{Exposure}_{H,L,s,5} \times (1 - \text{Friending Bias}_{H,L,s,1})$.

Bar 4: In this bar, instead of equating exposure, we equate bias, i.e., $\forall s$, set $\text{Friending Bias}_{H,L,s,1} = \text{Friending Bias}_{H,L,s,5}$, and plot $\sum_s \phi_{L,s,5} \times \text{Exposure}_{H,L,s,1} \times (1 - \text{Friending Bias}_{H,L,s,5})$.

Bar 5: In this bar, we report $EC_{H,L,5}$. This is equivalent to equating friending shares, exposure, and bias simultaneously, i.e., $\sum_s \phi_{L,s,5} \times \text{Exposure}_{H,L,s,5} \times (1 - \text{Friending Bias}_{H,L,s,5}) = EC_{H,L,5}$.

Figure 4: Figure 4 presents county-level and ZIP-level maps of exposure and bias for low-SES individuals. We define the following measures, calculated over all individuals whom we can assign to a setting. The differences relative to the measures in Figure 2 are that here we focus on $SES = L$, aggregate over the setting dimension s, and add a geographic dimension $a \in \{\text{county, ZIP}\}$. Let $N_{L,a}$ denote the number of low-SES individuals in area a (county or ZIP) and $a(i)$ denote i's area.

Panels A-B: These panels plot area-level (county- or ZIP-code-level) means of exposure based on the groups to which low-SES individuals belong. These area-level measures of the exposure of low-SES to high-SES individuals are computed by first calculating individuals' exposure as the sum of the product of their friend shares in each group $\phi_{i,g}$ and exposure in the individual's cohort-group, $\text{Exposure}_{H,g}^{c(i)}$, and then averaging over all low-SES individuals in an area a ($i \in L : a(i) = a$):

$$\text{Exposure}_{H,L,a} = \frac{1}{N_{L,a}} \sum_{\substack{i \in L \\ a(i)=a}} \sum_{g \in G(i)} \phi_{i,g} \times \text{Exposure}_{H,g}^{c(i)}. \tag{19}$$

Panels C–D: These panels plot area-level (county- or ZIP-code-level) means of friending bias based on the groups to which low-SES individuals belong and the friendships made in those groups. These area-level measures of friending bias are computed by first calculating individual-level friending bias as the sum of the product of their friend shares in each group $\phi_{i,g}$ and friending bias in that group, $\text{Friending Bias}_{H,i,g}$, and then averaging over all low-SES individuals in an area a ($i \in L : a(i) = a$):

$$\text{Friending Bias}_{H,L,a} = \sum_{\substack{i \in L \\ a(i)=a}} \sum_{g \in G(i)} \phi_{i,g} \times \text{Friending Bias}_{H,i,g}. \tag{20}$$

Figure 5: In Figure 5, we report exposure and bias estimates for colleges and high schools. These measures are obtained by averaging low-SES individuals' exposure and bias over all low-SES individuals in each group who make at least one friend in that group. Specifically, for a particular high school or college u[10]:

$$\text{Exposure}_{H,L,u} = \frac{1}{N_{L,u}} \sum_{i \in L} \sum_{u \in G(i)} \text{Exposure}_{H,u}^{c(i)} \tag{21}$$

and

$$\text{Friending Bias}_{H,L,u} = \frac{1}{N_{L,u}} \sum_{i \in L} \sum_{u \in G(i)} \text{Friending Bias}_{H,i,u}. \tag{22}$$

[10]These definitions of exposure and friending bias apply more generally to groups u in all settings $s \in \{$college, employer, high school, neighborhood, recreational group, religious group$\}$. For example, Table 1 presents a variance decomposition using these group-level measures and Table 5 reports summary statistics on exposure, friending bias, and economic connectedness for all settings. Supplementary Figure 1 analyzes friending bias across groups within different settings.

We can also compute the school- or college-level average economic connectedness to high-SES students among low-SES students. This measure is given as individuals' economic connectedness at that school, averaged over all low-SES individuals who attend the school and make at least one friend in that school:

$$\mathrm{EC}_{H,L,u} = \frac{1}{N_{L,u} w_H} \sum_{i \in L} \sum_{u \in G(i)} f_{H,i,u} \tag{23}$$

Figure 9: In Figure 9, we analyze the causal effects of being assigned to a high-school cohort with more high-SES peers on economic connectedness for low-SES students. We focus on connections between children with parents in the lowest and highest SES quintiles (rather than below vs. above median SES). Hence, we define different SES types $SES \in \{L', H'\}$, with L' denoting the set of bottom-quintile-SES individuals and H' denoting the set of top-quintile-SES individuals.

We first measure rates of exposure to high-SES students by school cohort. To do so, we first assign individuals to school cohorts based on their birth dates relative to the school entry cutoff date for their school (see Supplementary Table 2). We then define exposure in a given cohort c of high school u as $\mathrm{Exposure}^c_{H',u} = \frac{w^c_{H',u}}{w_{H'}}$, where $w^c_{H',u}$ is the share of top-quintile-SES people in cohort c, and $w_{H'} = 0.2$. Because these figures exploit cohort-level fluctuations in peers, we define exposure differently from how it is defined in Figure 2b: here, $w^c_{H',u}$ refers to exposure only *within* cohort c itself, rather than in cohorts $\{c-3, c-2, \ldots, c+3\}$.

We define the cohort-level deviation of exposure as mean exposure for the relevant cohort c in a given school u minus the mean for all other cohorts in the same school, weighting by the number of bottom-quintile-SES students in each cohort:

$$\Delta\mathrm{Exposure}^c_{H',u} = \mathrm{Exposure}^c_{H',u} - \sum_{j \neq c} p_{u,j} \times \mathrm{Exposure}^j_{H',u} \tag{24}$$

where $p_{u,j}$ is the share of bottom-quintile-SES students in cohort j out of all bottom-quintile-SES students in school u not in cohort c.

In Panel A, we examine the relationship between $\Delta\mathrm{Exposure}^c_{H',u}$ and an analogous cohort-level deviation measure for economic connectedness. To construct the latter, we start by defining $f_{H',i,u}$ as the fraction of top-quintile-SES friends that individual i makes in school u and cohort $c(i)$. Let $N_{L',u,c}$ denote the number of bottom-quintile-SES students in school u and cohort c. Then, the cohort-level deviation of economic connectedness is

$$\Delta\mathrm{EC}^c_{H',L',u} = \frac{1}{N_{L',u,c} w_{H'}} \sum_{\substack{i \in L' \\ c(i)=c}} f_{H',i,u} - \sum_{j \neq c} p_{u,j} \left(\frac{1}{N_{L',u,j} w_{H'}} \sum_{\substack{i \in L' \\ c(i)=j}} f_{H',i,u} \right). \tag{25}$$

Panel A presents a binned scatterplot of $\Delta\mathrm{EC}^c_{H',L',u}$ vs. $\Delta\mathrm{Exposure}^c_{H',u}$ and corresponding slope estimate from an OLS regression.

In Panel B, we construct a measure of friending bias using data from all cohorts excluding the focal cohort c as follows:

$$\text{Friending Bias}^{-c}_{H',L',u} = 1 - \frac{1}{N_{L',u,-c}} \sum_{\substack{i \in L' \\ c(i) \neq c}} \frac{f_{H',i,u}}{w^{c(i)}_{H',u}} \tag{26}$$

where $N_{L',u,-c}$ is the number of bottom-quintile-SES students in school u not in cohort c. We use this leave-out measure of bias to divide school-cohort cells into ten deciles. We then estimate OLS

regressions analogous to that in Panel A using the cohort-school cells in each of the ten deciles separately. Finally, we plot the slopes from these ten regressions vs. the mean level of (leave-out) friending bias in each decile.

B.6 Regression Discontinuity Design: Specifications and Robustness

In the main text, we use a regression discontinuity design to estimate the causal effect of exposure to high-SES peers on economic connectedness in high schools. Here, we discuss the estimating equations and identification assumptions underlying this design.

We begin from data at the individual level and define economic connectedness EC_i for individual i as five times the share of their high school friends in their cohort who have parental SES in the top quintile of the national parental SES distribution. We then consider every pair of adjacent cohorts, letting cohort 1 be the cohort with greater high-SES exposure and cohort 2 the other cohort. For each student in cohort 1, let d_i denote the number of days between their birthdate and the cutoff date between the two cohorts (using the cutoffs defined in Supplementary Table 2). For students in cohort 2, define the number of days between their birthdate and the cutoff between the two cohorts analogously, but code it to be negative. Each student in the two cohorts is then entered into the dataset with an accompanying d_i, positive for students in cohort 1 and negative for those in cohort 2.

Since we do this for every pair of cohorts, students will appear twice in the dataset (except for those in the first and last cohorts), once with d_i that is defined relative to the cutoff date defining the cohort younger than the student and once with d_i defined relative to the cutoff defining the cohort older than the student. Depending on the comparison of the student's cohort's own high-SES share relative to the preceding and following cohorts, these values of d_i could both be positive, negative, or be of different signs. For example, consider a student i in cohort $c(i)$. If i is born 15 days after the cutoff from cohort $c-1$ and the share of high-SES students in $i's$ school is lower in cohort c than $c-1$, then $d_i = -15$ for $(c-1, c)$ pair of cohorts. If the share of high-SES students is greater in cohort c than $c+1$, then $d_i = +350$ for the $(c, c+1)$ cohort pair.

We collapse the resulting individual-level dataset into cell means, defining cells by three dimensions: (1) quartile of changes in exposure across cohorts ($\triangle e_q$), (2) quartile of friending bias estimated using data from other cohorts excluding the two focal cohorts (b_q), and (3) days between individuals' dates of birth and school entry cutoff dates, defined as above (d).

Using data for a given set of exposure change by bias $\{\triangle e_q, b_q\}$ cells, we estimate regression specifications of the following form:

$$EC_{\triangle e_q, b_q, d} = \beta_0 + \beta_1 T_d + \beta_2 d + \beta_3 T_d \times d + \epsilon_{e_q, b_q, d}, \tag{27}$$

weighting by the number of individuals in each $\triangle e_q \times b_q \times d$ cell.

In equation (27), the outcome variable $EC_{\triangle e_q, b_q, d}$ is average EC in each $\triangle e_q \times b_q \times d$ cell. The indicator variable $T_d = 1\{d > 0\}$ takes a value of 1 when individuals fall on the side of the entry cutoff that has greater high-SES exposure. The coefficient of interest is β_1, which is an estimate of how much the economic connectedness of low-SES individuals changes in response to the jump in high-SES exposure. In Figure 10a, we focus on pairs of adjacent cohorts where the magnitude of the jump in the share of high-parental-SES students lies in top quartile of the distribution of changes in high-SES shares ($\triangle e_q = 4$); in Figure 10b, we report estimates for all four quartiles of the change in high-SES shares across cohorts. See the notes to Figure 10 for additional details on implementation of this specification.

The key assumption for (27) to yield an unbiased estimate of β_1 is that only high-SES exposure changes at the cutoff and that individuals on either side of the cutoff are similar on all other relevant characteristics ϵ that may affect EC. To evaluate this identification assumption, we assess whether observable characteristics are balanced on the two sides of the cutoff. In Supplementary Figure 7,

we replicate Figure 10a replacing EC_i with individual i's total number of friends and sex as the outcome variables. Both the number of friends and share female trend smoothly around the cutoff, supporting the validity of the identification assumption.

In the baseline regression specification used in Figure 10a, we include students within 200 days of the entry cutoff ($-200 \leq d_i \leq 200$). To evaluate the sensitivity of our estimates, we consider different choices of bandwidth. Supplementary Figure 8 presents histograms of estimates of β_1 for low-bias and high-bias schools for bandwidths ranging from 20 to 300 days in 10 day increments. The estimates are all clustered around the baseline estimates of 0.39 for low-bias schools and 0.33 for high-bias schools, showing that our conclusions are robust to the choice of the bandwidth.

C Supplementary Discussion

C.1 Level of Aggregation and Exposure vs. Friending Bias

As we discuss in the main text, the distinction between friending bias and exposure depends on the level at which one measures exposure. For example, when measuring exposure and bias at the school level, tracking to different classes may result in higher observed friending bias at the school level; however, that friending bias might be the result of lower exposure (lack of integration by SES) at the classroom level. Similar phenomena arise in other settings: for instance, ZIP codes that exhibit high friending bias may be highly segregated by income across blocks within the ZIP code. Indeed, we find that counties with higher levels of residential income segregation have higher levels of mean friending bias, even after controlling for the exposure of low-SES individuals to high-SES individuals in their groups, suggesting that friending bias is partly the result of residential income segregation within ZIP codes.

Despite these observations, the efficacy of interventions in increasing economic connectedness depends on the importance of exposure and friending bias at the units of aggregation we analyze, since those are often units of interest for policy makers (e.g., high schools, colleges, and workplaces). For example, if low-SES students have few high-SES friends in their high schools because of a lack of exposure at the high school level, then integration across schools could be effective in increasing EC. But if low-SES students have few high-SES friends even in schools that are socioeconomically diverse, other within-school interventions (such as changes in tracking) may be necessary to increase EC.

C.2 Relationship Between Exposure and Average Incomes Across Areas

In the main text, we show that mean levels of high-SES exposure and the overall share of high-SES people in a given ZIP code or county are highly correlated with each other. Although these correlations are intuitive, they are not mechanical because exposure for low-SES individuals is constructed at the group level, based on, e.g., the specific high schools or recreational groups in which low-SES people participate.

Differences between exposure and the local share of high-SES (or high-income) people provide new information about the degree to which groups in an area are integrated by SES. In particular, two ZIP codes that have similar shares of high-SES people overall but different levels of high-SES exposure must differ in their degree of income segregation across groups; that is, the extent to which low- and high-SES residents attend the same schools, work at the same employers, participate in the same recreational groups, etc. Relatedly, note that high-SES and low-SES individuals who live in the same ZIP code may have different levels of high-SES exposure because they may be members of different groups within the same area.

C.3 Relationship Between Upward Mobility and Friending Bias

Our finding that upward mobility is strongly related to friending bias echoes the findings in Figure 11 and Table 5a of Chetty et al. (2022), where we showed that economic connectedness strongly predicts economic mobility even conditional on income levels in the areas where low-SES people live. Here, we further establish that controlling for SES in the specific *groups* that low-SES individuals belong to (i.e., their high schools, religious groups, etc.)—which differs from overall SES in the areas they live in for the reasons discussed in the preceding subsection—does not affect the relationship between economic connectedness and upward mobility.

C.4 Friending Bias at Evanston Township High School: Ethnographic Evidence

The high degree of friending bias at Evanston Township High School (ETHS) observed in our data mirrors historical ethnographic accounts of racial and socioeconomic segregation at the school (Barr 2014). It is also consistent with more recent discussions within the student body. For instance, the student newspaper *The Evanstonian* has frequently reported on racial and socioeconomic disconnections within the student body. In one newspaper article, Jacobs and Martinez-Olsen (2019) highlight socioeconomic disparities in access to extra-curricular activities, arguing that the high cost of prom tickets prevents students from low-income families from attending: "(Prom) is a once in a lifetime opportunity so you'll go all out. It's said to be "the time of your life" but not all students have access to that kind of money."

Another article in *The Evanstonian* describes the efforts of a student group called the Paw Patrol—which seeks to foster school spirit by organizing student cheering at high school sporting events—to increase across-group interaction at ETHS (Dain et al. 2021): "In terms of diversity, Paw Patrol has been successful in getting a range of students involved. [...] However, diversity and integration are two commonly confused terms with entirely separate meanings. Diversity applies solely to demographics, whereas integration investigates the interpersonal connections across those demographics. Paw Patrol has undergone improvement regarding diversity but still requires growth in terms of integration. 'I definitely feel like there is segregation, and we've talked about it,' [student leader Maya] Wallace explains."

References

Abbiasov, Timur (2021). "The Contribution of Urban Parks to Social Ties between Neighborhoods". *Working Paper*.

Alan, Sule, Enes Duysak, Elif Kubilay, and Ipek Mumcu (2020). "Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice". *The Review of Economics and Statistics*, pp. 1–45.

Alesina, Alberto, Reza Baqir, and William Easterly (1999). "Public Goods and Ethnic Divisions". *The Quarterly Journal of Economics* 114.4, pp. 1243–1284.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew (1954). *The nature of prejudice*. Addison-Wesley.

Athey, Susan, Billy A Ferguson, Matthew Gentzkow, and Tobias Schmidt (2020). *Experienced Segregation*. Working Paper 27572. National Bureau of Economic Research.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2018a). "Social Connectedness: Measurement, Determinants, and Effects". *Journal of Economic Perspectives* 32.3, pp. 259–280.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel (2018b). "The Economic Effects of Social Networks: Evidence from the Housing Market". *Journal of Political Economy* 126.6, pp. 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel (2019). "House Price Beliefs And Mortgage Leverage Choice". *The Review of Economic Studies* 86.6, pp. 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel (2020a). "Social Connectedness in Urban Areas". *Journal of Urban Economics* 118, p. 103264.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert J Richmond, and Johannes Stroebel (2020b). *International Trade and Social Connectedness*. Working Paper 26960. National Bureau of Economic Research.

Barr, Mary (2014). *Friends Disappear: The Battle for Racial Equality in Evanston*. University of Chicago Press.

Blau, Peter M. (1977). "A Macrosociological Theory of Social Structure". *American Journal of Sociology* 12.1, pp. 26–54.

Blau, Peter M. and Joseph E. Schwartz (1984). *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations*. Academic Press.

Boston Voyager (2018). *Meet Jon Feinman of InnerCity Weightlifting in Dorchester and Kendall Square*.

Bush, Melodye and Kyle Zinth (2011). *Kindergarten Entrance Ages: A 35 Year Trend Analysis*.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West (2013). "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation". *Econometrica* 81.3, pp. 855–882. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10168.

Cheng, Siwei and Yu Xie (2013). "Structural effect of size on interracial friendship". *Proceedings of the National Academy of Sciences* 110.18, pp. 7165–7169.

Chetty, Raj and John Friedman (2019). "A Practical Method to Reduce Privacy Loss When Disclosing Statistics Based on Small Samples". *American Economic Review Papers and Proceedings* 105.5, pp. 414–420.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2018). *The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility*. Working Paper 25147. National Bureau of Economic Research.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2020). "Income Segregation and Intergenerational Mobility Across Colleges in the United States". *Quarterly Journal of Economics* 135.3, pp. 1567–1633.

Chetty, Raj and Nathaniel Hendren (2018). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates". *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

Chetty, Raj, M Stepner, S Abraham, S Lin, B Scuderi, N Turner, A Bergeron, and D Cutler (2016). "The Association between Income and Life Expectancy in the United States, 2001-2014". *Journal of the American Medical Association* 315.16, pp. 1750–1766. eprint: /data/journals/jama/935237/jsc160006.pdf.

Chetty, Raj et al. (2022). "Social Capital I: Measurement and Associations with Economic Mobility". *NBER Working Paper*.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin (2009). "An Economic Model of Friendship: Homophily, Minorities, and Segregation". *Econometrica* 77.4, pp. 1003–1045.

— (2010). "Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation". *Proceedings of the National Academy of Sciences* 107.11, pp. 4857–4861.

Dain, Lauren, Jessica Sehgal, Meg Houseworth, Bridget Baker, Stella Israelite, Sophie Milner-Gorvine, and Zoe Kaufman (2021). "ETHS activities and spaces grapple with inequities". *The Evanstonian*.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales (2019). "How segregated is urban consumption?" *Journal of Political Economy* 127.4, pp. 1684–1738.

Dong, Xiaowen, Alfredo J. Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland (2020). "Segregated Interactions in Urban and Online Space". *EPJ Data Science* 9.1, p. 20.

Doob, Christopher B (2019). *Social inequality and social stratification in US society*. Routledge.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith (2006). "Calibrating Noise to Sensitivity Analysis". Halevi, Shai and Tal Rabin. *Theory of Cryptography*. Springer Berlin Heidelberg, pp. 265–284.

Elder, Todd E. and Darren H. Lubotsky (2009). "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers". *Journal of Human Resources* 44.3, pp. 641–683.

Feinman, Jonathan (2022). *A Letter from our Founder*. URL: https://www.innercityweightlifting.org/founderletter.

Feld, Scott L. (1982). "Social Structural Determinants of Similarity among Associates". *American Sociological Review* 47.6, pp. 797–801.

Fischer, Claude S and Greggor Mattson (2009). "Is America Fragmenting?" *Annual Review of Sociology* 35, pp. 435–455.

Gee, Laura K. (2018). "The More You Know: Information Effects on Job Application Rates in a Large Field Experiment". *Management Science* 65.5, pp. 2077–2094.

Gee, Laura K., Jason Jones, and Moira Burke (2017). "Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network". *Journal of Labor Economics* 35.2, pp. 485–518.

Gentzkow, Matthew and Jesse M Shapiro (2011). "Ideological segregation online and offline". *The Quarterly Journal of Economics* 126.4, pp. 1799–1839.

HKS (2019). *Thinking Outside the Checkbox: How HKS Teamed Up With a Dallas School to Connect Divided Campus*. URL: https://www.hksinc.com/our-news/articles/thinking-outside-the-checkbox-how-hks-teamed-up-with-a-dallas-school-to-connect-divided-campus/ (visited on 11/26/2021).

Jacobs, Caroline and Aldric Martinez-Olsen (2019). "Racial and socioeconomic implications in how ETHS lost its spirit". *The Evanstonian*.

Jeffres, Leo W., Cheryl C. Bracken, Guowei Jian, and Mary F. Casey (2009). "The Impact of Third Places on Community Quality of Life". *Applied Research in Quality of Life* 4.4, p. 333.

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H. Fowler (2013). "Inferring Tie Strength from Online Directed Behavior". *PLOS ONE* 8.1, e52168.

Kim, Alicia (2018). "Berkeley High School tackles ethnic disparities on campus". *The Daily Californian*.

Klinenberg, Eric (2018). *Palaces for the people: How social infrastructure can help fight inequality, polarization, and the decline of civic life*. Crown.

Levy, Brian L., Nolan E. Phillips, and Robert J. Sampson (2020). "Triple Disadvantage: Neighborhood Networks of Everyday Urban Mobility and Violence in U.S. Cities". *American Sociological Review* 85.6, pp. 925–956. eprint: https://doi.org/10.1177/0003122420972323.

Oldenburg, Ray (1999). *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. 3rd edition. New York : Berkeley, Calif.: Marlowe & Company.

Owens, Ann, Sean F. Reardon, and Christopher Jencks (2016). "Income Segregation Between Schools and School Districts". *American Educational Research Journal* 53.4, pp. 1159–1197.

Pallais, Amanda and Jenna Anders (2021). "Why Can't We Be Friends?: Theory and Evidence on Friendship Formation". *Harvard University Working Paper*.

Pew Research Center (2014). *Religious Landscape Study*. Tech. rep.

Putnam (2016). *Our kids: The American dream in crisis*. Simon and Schuster.

— (2020). *The Upswing: How American Came Together a Century Ago and how We Can Do it Again*. Simon and Schuster.

Reardon, Sean and Kendra Bischoff (2011). "Income Inequality and Income Segregation". *American Journal of Sociology 116 (4): 1092-1153*.

Reardon, Sean and Ann Owens (2014). "60 years after Brown: Trends and consequences of school segregation". *Annual Review of Sociology* 40, pp. 199–218.

Reme, Bjørn-Atle, Andreas Kotsadam, Johannes Bjelland, Pål Roe Sundsøy, and Jo Thori Lind (2022). "Quantifying Social Segregation in Large-Scale Networks". *Scientific Reports* 12.1, pp. 1–7.

Rohrer, Julia M., Tamas Keller, and Felix Elwert (2021). "Proximity can induce diverse friendships: A large randomized classroom experiment". *PLOS One* 16.8, e0255097.

Sacerdote, Bruce (2011). "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 3. Elsevier, pp. 249–277.

Smith, Jeffrey A, Miller McPherson, and Lynn Smith-Lovin (2014). "Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004". *American Sociological Review* 79.3, pp. 432–456.

Terry, Leigh (2014). *Greek diversity: Why not socioeconomic too?* URL: https://cw.ua.edu/17926/opinion/greek-diversity-why-not-socioeconomic-too/ (visited on 11/26/2021).

U.S. Census Bureau (2017). *2012–2016 American Community Survey 5-Year Estimates*. Tech. rep.

Villano, Kenton J. (2011). "Equally Unprotected: Assessing Whether Evanston Township High School's Former or Future Freshman Honors Curriculum Runs the Risk of Violating Students' Rights". *The Loyola University Chicago Childlaw and Eduction Institute Forum* 2011.36.

Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson (2018). "Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities". *Proceedings of the National Academy of Sciences* 115.30, pp. 7735–7740.

White, Karl R (1982a). "The relation between socioeconomic status and academic achievement." *Psychological Bulletin* 91.3, p. 461.

— (1982b). "The relation between socioeconomic status and academic achievement." *Psychological Bulletin* 91.3, pp. 461–481.

Wilson, Riley (2020). "The Impact of Social Networks on EITC Claiming Behavior". *The Review of Economics and Statistics*, pp. 1–45.

Wuthnow, Robert (2002). "Religious involvement and status-bridging social capital". *Journal for the Scientific Study of Religion* 41.4, pp. 669–684.

Zeng, Zhen and Yu Xie (2008). "A preference-opportunity-choice framework with applications to intergroup friendship." *American Journal of Sociology* 114.3, pp. 615–48.

TABLE 1: EC, Exposure, and Friending Bias: Variation across Areas and Settings

A. Explaining Across-Group Level Variation in EC, Exposure, and Bias: Counties vs. Settings

	EC (1)	Exposure (2)	Friending Bias (3)
Share of Variation Across Counties	0.24	0.30	0.04
Share of Variation Across Settings	0.14	0.09	0.13

B. Explaining Across-ZIP Code Variation in EC, Exposure, and Bias

	EC (1)	Exposure (2)	Friending Bias (3)
Standard Deviation Across ZIP Codes	0.20	0.25	0.09
Share of ZIP-Code-Level Variation Across Counties	0.45	0.43	0.24

Notes: Panel A shows how much of the group-level variation in EC, exposure, and friending bias is between counties vs. settings. The first column reports the adjusted R^2 from two separate regressions: one where mean economic connectedness (EC) in each group is regressed on county indicators (a group's county is defined as the modal residential county of individuals assigned to that group), and another where EC is regressed on the six setting indicators. The unit of observation in these regressions is a single group such as a particular high school, neighborhood, or recreational group. EC is defined as twice the average share of above-median-SES friends in the group among below-median-SES people within each group. Columns 2 and 3 replicate column 1 with exposure and friending bias as the dependent variables. Exposure is defined as twice the share of above-median-SES individuals in the group. Friending bias is defined as one minus the mean ratio of the share of above-median-SES friends to the share of above-median-SES peers in the group, averaging over below-median-SES individuals in the group (see Supplementary Information B.5). Panel B focuses on the variation in mean EC, exposure, and friending bias at the ZIP code level. The first row reports the standard deviation of EC, exposure, and bias across ZIP codes. The second row shows the adjusted R^2s from regressions of ZIP-code level means of these three variables on county indicators. To construct ZIP code level means, we aggregate individual-level measures of EC, exposure, and friending bias as described in the notes to Figure 4 and Supplementary Information B.5. All statistics are weighted by the number of below-median-SES individuals in each group (Panel A) or ZIP code (Panel B).

TABLE 2: Associations between Friending Bias, Exposure and Upward Income Mobility across Areas

Dependent Variable:	log(Upward Mobility)						log(Causal Upward Income Mobility)
	ZIP Codes				Counties		Counties
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log (Economic Connectedness)	0.236***		0.227***		0.272***		
	(0.01)		(0.01)		(0.02)		
log (High-SES Exposure)		0.248***		0.224***		0.286***	0.116***
		(0.01)		(0.02)		(0.02)	(0.02)
log (1 - Friending Bias)		0.185***		0.236***		0.142*	0.339***
		(0.03)		(0.04)		(0.08)	(0.07)
County FEs	No	No	Yes	Yes	No	No	No
Observations	24,200	24,200	24,200	24,200	2,986	2,986	2,136
R-squared	0.42	0.43	0.71	0.71	0.38	0.39	0.03

Notes: This table presents estimates from OLS regressions of log upward income mobility on log economic connectedness (EC) and other covariates. The coefficients can be interpreted as the elasticity of upward mobility with respect to the relevant covariate. In columns 1-6, upward income mobility is obtained from the observational measures in the Opportunity Atlas (Chetty et al. 2018), and is defined as the predicted household income rank in adulthood for children with parents at the 25th percentile of the national income distribution. Columns 1-4 present regressions at the ZIP code level. In column 1, the only independent variable is log economic connectedness, defined here as the product of mean high-SES exposure and 1 - mean friending bias in the ZIP code; see Supplementary Information B.5. In column 2, the independent variables are the log of mean high-SES exposure and the log of 1 - mean friending bias; see Supplementary Information B.5. We start from individual-level statistics to compute ZIP code-level and county-level means of exposure and friending bias (see Supplementary Information B.5). At the individual level, exposure is defined as the weighted average of two times the fraction of above-median-SES members of the groups in which a below-median-SES individual participates, weighting each group by the individual's share of friends in that group. Friending bias is defined as one minus the weighted average of the ratio of the share of high-SES friends to the share of high-SES peers in the groups in which a low-SES individual participates, again weighting each group by the individual's share of friends in that group. Columns 3 and 4 replicate columns 1 and 2 adding county fixed effects. Columns 5 and 6 replicate columns 1 and 2 at the county level instead of ZIP code level. Column 7 replicates column 6 using counties' causal effects on upward mobility, defined as the mean predicted household income rank in adulthood for children with parents at the 25th percentile of the income distribution overall in the U.S. plus 20 times the raw annual causal exposure effect of growing up in the county reported in Chetty and Hendren 2018. Regressions in columns 1-6 are weighted by the number of individuals in the primary analysis sample with below-median SES in the county or ZIP code. The regression in column 7 is weighted by the inverse of the squared standard error of the estimated annual causal exposure effect of growing up in that county reported in Chetty and Hendren 2018. Standard errors (reported in parentheses) are clustered at the commuting zone level for county-level regressions and at the county level for ZIP code-level regressions. Asterisks indicate the level of significance: *10%, **5%, ***1%.

TABLE 3: Validation of SES Predictions using Publicly Available Data

Setting	Benchmark	Facebook SES Measure	Correlation with % Above-Median SES in Facebook Data
ZIP Codes	% of individuals with household income above the national median (ACS)	Own SES	0.88
High Schools	% of students not eligible for free or reduced lunch (NCES)	Parental SES	0.85
Colleges	% of students with parental household income in the top two quintiles of the national distribution (tax data)	Parental SES	0.91

Notes: This table evaluates the accuracy of the SES measures that constructed in the Facebook data as well as the methods used to assign individuals to specific groups (i.e., specific neighborhoods, high schools, and colleges). Concretely, we compare the fraction of above-median-SES individuals in each ZIP code, high school, and college to estimates of the fraction of high-income members of these groups from publicly available administrative data sources. In the first row, we correlate the fraction of households with above-median household income within each ZIP code, as calculated by Chetty et al. 2018 using data from the 2014-2018 American Community Survey, with the estimated proportion of Facebook users in our primary sample with above-median SES. This correlation is weighted by ZIP code population based on the 2018 ACS. In the second row, for each public high school, we calculate the 5-year median of the fraction of students who are not eligible for free or reduced price lunch (based on the 2014-2018 NCES Common Core of Data). We correlate this measure with the fraction of students with above-median parental SES in the Facebook data, using individuals born between 2000 and 2004 to match the cohorts observed in the NCES data. This correlation is weighted by the number of students in grades 9 to 12, as reported in the NCES data. In the third row, we correlate the fraction of students with parental income in the top two quintiles of the national distribution in each college, as calculated by Chetty et al. 2020 using tax records, with our corresponding estimates of the proportion of students with above-median parental SES from the Facebook data. This correlation is weighted by the number of students in the relevant college cohort. See Supplementary Information A for further details on the publicly available data sources.

TABLE 4: Summary Statistics for Analysis Samples

A. Primary Analysis Sample

	Mean (1)	SD (2)	P10 (3)	P25 (4)	P50 (5)	P75 (6)	P90 (7)	ACS Mean (8)	ACS SD (9)
Age (Years)	33.9	7.6	26	29	33	38	42	34.2	5.8
Male	46.3%	49.9%						50.2%	50.0%
English Language	93.5%	24.7%						94.4%	23.0%
Years on Facebook	11.4	4.0	4.4	10.0	12.5	13.9	15.5		
Number of Facebook Friends	579.3	603.6	100	191	393	751	1,267		
Share of Friendships Assigned to a Setting	30.3%	20.3%	3.1%	12.6%	29.6%	45.8%	58.3%		
College Graduation Rate (County-Level)	57.0%	15.2%	37.7%	46.7%	55.4%	67.0%	77.4%	58.1%	15.7%
Median Household Income ($, County-Level)	57,788	15,100	42,043	47,141	55,293	64,422	78,800	57,980	15,858
Share White (County-Level)	60.3%	21.7%	29.6%	43.4%	62.0%	78.5%	88.7%	58.8%	22.7%
N = 70.3 million Facebook users									

B. Subsample with Non-Missing Parents and High Schools

	Mean (1)	SD (2)	P10 (3)	P25 (4)	P50 (5)	P75 (6)	P90 (7)	ACS Mean (8)	ACS SD (9)
Age (Years)	31.6	5.8	25	27	31	35	39	34.2	5.8
Male	48.5%	50.0%						50.2%	50.0%
English Language	99.1%	9.5%						94.4%	23.0%
Years on Facebook	12.4	3.1	8.6	11.7	12.9	14.2	15.5		
Number of Facebook Friends	699.5	602.3	172	301	539	903	1,399		
Share of Friendships Assigned to a Setting	43.0%	18.5%	17.6%	29.1%	43.4%	57.0%	67.5%		
College Graduation Rate (County-Level)	57.0%	15.2%	37.7%	46.7%	55.4%	67.0%	77.4%	58.1%	15.7%
Median Household Income ($, County-Level)	56,911	14,410	41,908	46,859	54,469	63,108	77,627	57,980	15,858
Share White (County-Level)	64.6%	21.0%	32.7%	48.9%	67.8%	82.5%	90.8%	58.8%	22.7%
N = 19.4 million Facebook users									

Notes: Panel A presents summary statistics for our primary analysis sample, which consists of individuals between the ages of 25 and 44 as of May 28, 2022, who reside in the United States, have been active on the Facebook platform at least once in the previous 30 days, have at least 100 U.S.-based Facebook friends, have a non-missing residential ZIP code, and for whom we are able to allocate at least one friend to a setting using the algorithm described in Section VI.B. Panel B replicates Panel A for the sample used to measure childhood economic connectedness—individuals in the primary analysis sample whom we can link to parents with valid SES predictions (see Supplementary Information A.2 in Chetty et al. 2022) and can be assigned a high school (see Supplementary Information B.1). For each variable, columns 1-7 present means, standard deviations, and selected percentiles. Columns 8 and 9 report means and standard deviations for the corresponding variables using the nationally representative 2012-2016 American Community Survey (ACS) for median household income and the 2014-2018 ACS for the remaining variables. Age and gender are self-reported by individuals. *English Language* refers to whether users set their language to English in the Facebook data and to the share of individuals who speak English only or speak English "very well" in the ACS data. *Years on Facebook* and *Number of Facebook Friends* are only observed in the Facebook data. In both panels, we report the number of Facebook friends within our primary analysis sample, as opposed to the total number of Facebook friends. *Share of Friends Assigned to a Setting* refers to the set of friends that we are able to assign to one of the six settings, as described in Supplementary Information B.2. The remaining variables are obtained from the 2014–2018 American Community Survey (ACS) and are assigned based on an individual's residential county. The ACS means for these variables are population-weighted averages of the same county-level variables.

TABLE 5: Friend Shares, Economic Connectedness, Exposure, and Bias by Setting and SES

Setting	Low-SES Individuals				High-SES Individuals			
	Share of Friends Made in Setting	Economic Connectedness	High-SES Exposure	Friending Bias	Share of Friends Made in Setting	Economic Connectedness	High-SES Exposure	Friending Bias
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High School	0.467	0.890	0.935	0.043	0.533	1.446	1.246	-0.200
College	0.068	1.168	1.223	0.047	0.211	1.651	1.515	-0.102
Employer	0.016	0.915	0.961	0.054	0.019	1.484	1.334	-0.133
Recreational Group	0.017	0.888	0.924	0.025	0.018	1.463	1.395	-0.066
Religious Group	0.019	0.869	0.864	-0.032	0.016	1.450	1.300	-0.147
Neighborhood	0.413	0.680	0.820	0.166	0.202	1.428	1.174	-0.282

Notes: This table shows the share of friends made in each of the six settings we consider along with mean levels of economic connectedness, high-SES exposure, and friending bias in each setting. Columns 1-4 report these statistics for low-SES individuals and columns 5-8 report these statistics for high-SES individuals. See the notes to Figure 2 and the description in Supplementary Information B.5 for definitions of each of these measures. In columns 1 and 5, we report average normalized friend shares — each individual's raw friend share in a setting divided by the sum of their raw friend shares across the six settings — because the raw friend shares can sum to more than 1 due to the double counting of friendships assigned to multiple settings (see Section VI.B). Note that because exposure and bias covary across groups within settings, mean economic connectedness for each setting is not identical to the product of mean exposure and one minus mean friending bias at the setting level.

SUPPLEMENTARY TABLE 1: EC, Exposure, and Bias Across High Schools and Colleges Using Parental vs. Own SES Ranks

	Economic Connectedness		Exposure		Friending Bias	
Variables	High School	College	High School	College	High School	College
	(1)	(2)	(3)	(4)	(5)	(6)
Corr. Using Par vs Own Rank	0.83	0.84	0.84	0.86	0.59	0.61
Mean Using Parent SES Rank	0.92	1.11	0.93	1.13	0.01	0.02
SD Using Parent SES Rank	0.31	0.28	0.31	0.28	0.04	0.03
Mean Using Own SES Rank	0.92	1.22	0.96	1.27	0.03	0.04
SD Using Own SES Rank	0.33	0.36	0.34	0.35	0.06	0.06

Notes: This table compares estimates of economic connectedness, high-SES exposure, and friending bias for high schools and colleges computed using individuals' own SES ranks in adulthood vs. parental SES ranks. We calculate EC, exposure, and friending bias within schools and colleges as described by equations 21, 22 and 23 in Supplementary Information B.5. We restrict the sample to the schools and colleges shown in Figure 5. The first row shows the weighted correlation of the relevant statistics across schools/colleges when computed using parental vs. own SES, weighting by the number of below-median-parental-SES students in each high school/college born between 1990 and 2000. The second and third rows show the mean and standard deviation of the relevant statistic across schools/colleges computed using parental SES rank; the fourth and fifth rows replicate these statistics using own SES rank. When using parental SES rank, we average over below-median-SES individuals born between 1990 and 2000 whom we can link to parents because parental linkage rates are highest for those cohorts. When using own SES rank, we use all below-median-SES individuals born between 1986 and 1996.

SUPPLEMENTARY TABLE 2: School Entrance Cutoffs by State and Birth Cohort

State	Birth Cohorts	Cutoff Date	State	Birth Cohorts	Cutoff Date
AL	1976–1985	1-Oct	MO	1976–1979	1-Oct
	1986–1998	1-Sep		1982–1991	1-Jul
AK		N/A		1992–1998	1-Aug
AZ	1987–1998	1-Sep	MT	1985–1998	10-Sep
AR	1976–1992	1-Oct	NE	1976–1998	15-Oct
	1993–1998	15-Sep	NV	1979–1998	30-Sep
CA	1976–1998	2-Dec	NH	1988–1995	1-Oct
CO	1976–1998	15-Sep	NJ	1979–1998	1-Oct
CT	1976–1998	30-Sep	NM	1976–1998	1-Sep
DC		N/A	NY	1976–1998	1-Dec
DE	1980–1987	31-Dec	NC	1976–1998	15-Oct
	1991–1998	31-Aug	ND	1976–1989	31-Aug
FL	1979–1998	1-Sep	OH	1976–1998	30-Sep
GA	1976–1998	1-Sep	OK	1976–1998	1-Sep
HI	1976–1998	31-Dec	OR	1976–1979	15-Nov
ID	1976–1984	1-Oct		1981–1998	1-Sep
	1988–1998	1-Sep	PA	1983–1998	1-Sep
IL	1976–1980	1-Dec	RI	1976–1998	31-Dec
	1983–1998	1-Sep	SC	1976–1989	1-Nov
IN	1980–1984	1-Sep		1990–1998	1-Sep
	1987–1992	1-Jun	SD	1976–1980	1-Sep
IA	1976–1998	15-Sep	TN	1976–1978	31-Oct
KS	1976–1998	1-Sep		1980–1998	30-Sep
KY	1976–1998	1-Oct	TX	1976–1998	1-Sep
LA	1976–1980	31-Dec	UT	1976–1977	1-Nov
	1984–1998	30-Sep		1978–1998	2-Sep
ME	1976–1998	15-Oct	VT	1976–1985	31-Dec
MD	1976–1997	31-Dec	VA	1976–1980	31-Dec
MA	1980–1998	1-Sep		1988–1998	30-Sep
MI	1976–1998	1-Dec	WA	1982–1998	31-Aug
MN	1976–1998	1-Sep	WV	1976–1978	1-Nov
MS	1976–1998	1-Sep		1980–1998	31-Aug
			WI	1976–1998	1-Sep
			WY	1976–1980	15-Sep

Notes: This table reports the school entrance cutoff dates by state and year that we use for the quasi-experimental cross-cohort and regression discontinuity analyses in Figures 9 and Figure 10. We obtain these cutoffs from Elder and Lubotsky (2009) and Bush and Zinth (2011). The cutoffs determine whether children are assigned to earlier or later school cohorts based on their dates of birth.

SUPPLEMENTARY TABLE 3: Summary Statistics on Group and Friendship Assignment

A. Summary Statistics on Assignment to Groups

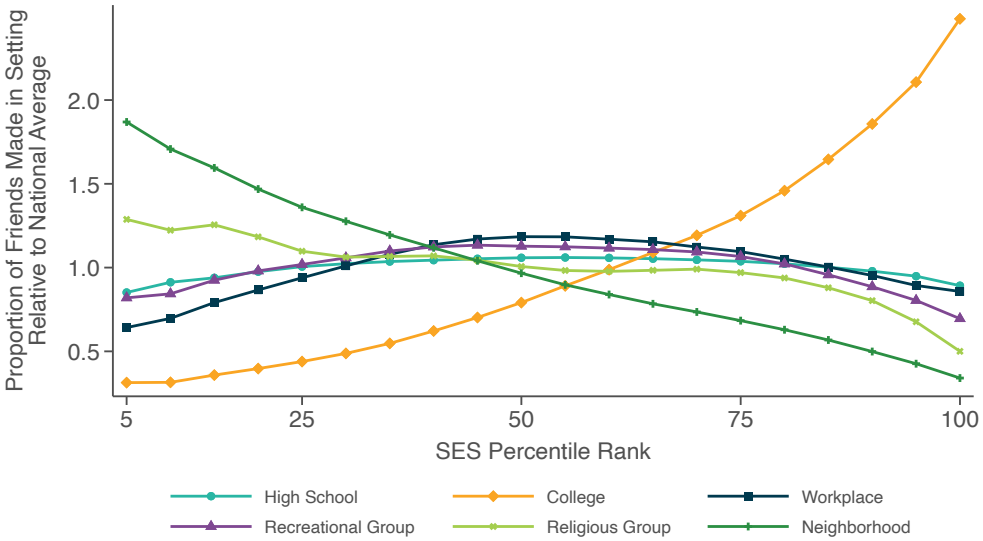
	Share Users Assigned			Share with > 1 Group (conditional on assignment)	Mean Users per Group	S.D. of Users per Group	Number of Groups
	All	Low-SES	High-SES				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Religious Group	17.9%	16.0%	19.8%	38.4%	39	64	332,042
College	42.9%	24.5%	61.3%	17.8%	530	821	2,921
Employer	20.9%	17.0%	24.7%	9.5%	45	881	334,388
High School	74.9%	67.0%	82.8%	3.3%	115	134	24,322
Neighborhood	100.0%	100.0%	100.0%	0.0%	2,484	3,750	29,062
Rec. Group	29.8%	22.7%	37.0%	45.9%	17	48	1,291,559

B. Overlap of Friendship Assignments Across Settings

	Rel. Group	College	Employer	High School	Neighborhood	Rec. Group	Uniquely Assigned
Religious Group	100.0%	1.8%	0.3%	12.0%	33.2%	3.8%	48.8%
College	0.3%	100.0%	0.4%	27.6%	7.7%	0.7%	63.3%
Employer	0.4%	2.4%	100.0%	5.0%	14.2%	1.0%	77.1%
High School	0.3%	3.8%	0.1%	100.0%	12.8%	0.7%	82.3%
Neighborhood	1.7%	2.1%	0.6%	25.4%	100.0%	2.0%	68.2%
Recreational Group	3.5%	3.2%	0.7%	25.6%	36.1%	100.0%	30.9%

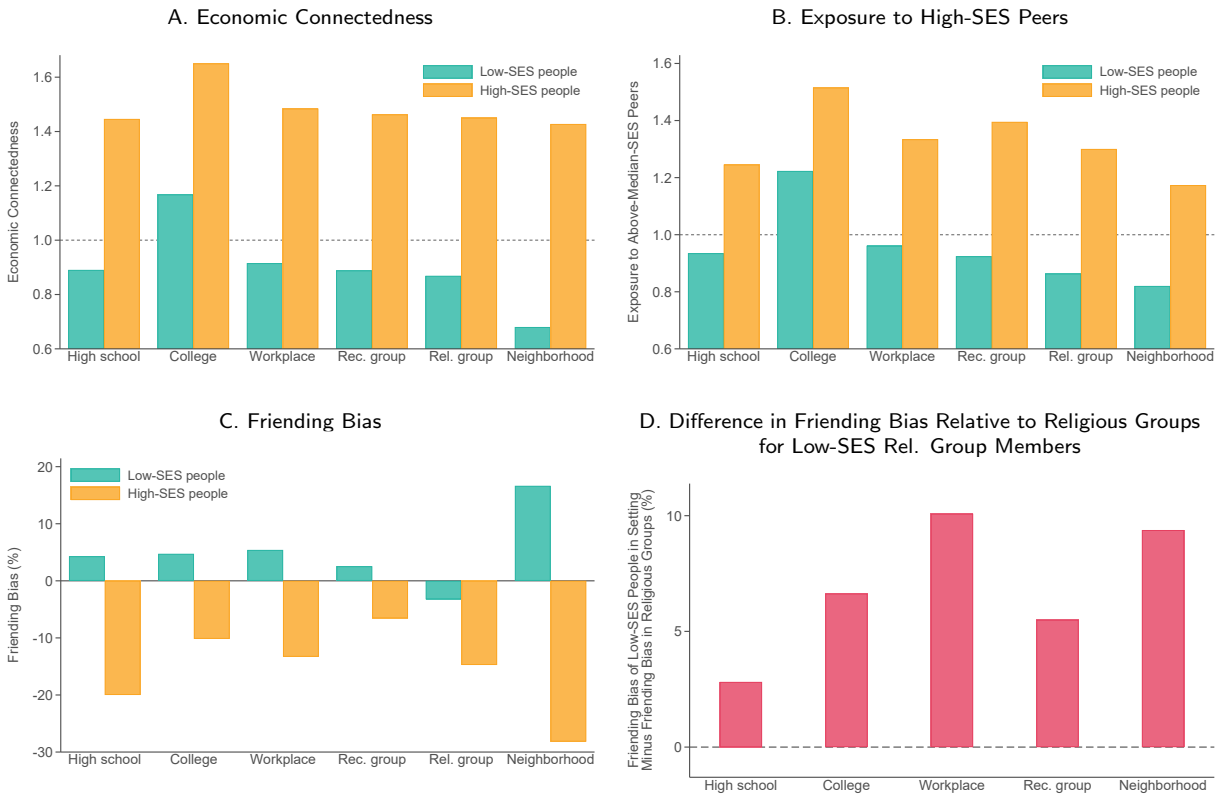
Notes: Panel A shows summary statistics regarding the assignment of individuals to each of the six settings (shown in the rows) that we analyze. The sample used in this table includes all individuals between the ages of 25 and 44 as of May 28, 2022 who reside in the United States, have been active on the Facebook platform at least once in the previous 30 days, have at least 100 U.S.-based Facebook friends, have a non-missing residential ZIP code, and for whom we are able to allocate at least one friend to a setting using the algorithm described in Section VI.B. Column 1 lists the share of individuals in our primary analysis sample assigned to one or more groups in the relevant setting (e.g., the fraction of individuals assigned to a religious group in row 1). Columns 2 and 3 replicate column 1 for low-SES and high-SES individuals, respectively. Column 4 shows the share of users assigned to more than one group given that they are assigned to at least one group in a setting. As an example, for religious groups, this is the share of users assigned to more than one religious group conditional on being assigned at least one religious group. Columns 5 and 6 show means and standard deviations of group size (Facebook users in the analysis sample) by setting; for high schools and colleges, we report the number of users per cohort. Column 7 shows the number of unique groups within each setting. Panel B reports statistics that characterize the share of friendships allocated to multiple settings. The off-diagonal elements show the fraction of friendships made in the setting shown in a given row that are also assigned to the setting shown in the column; for instance, 1.8% of the friendships assigned to religious groups are also assigned to colleges. The diagonal elements correspond to own-group pairs and are thus 100%. The last column shows a lower bound on the fraction of friendships that can be unambiguously assigned to a given group type. This bound is calculated by subtracting the off-diagonal elements in each row from 100%; intuitively, this bound assumes that all intersections of three or more groups are empty.

FIGURE 1: Friending Rates by Setting and Socioeconomic Status



Notes: This figure shows how friending rates vary across settings by the socioeconomic status (SES) percentile rank of individuals in our primary analysis sample. The primary analysis sample consists of individuals between the ages of 25 and 44 as of May 28, 2022 who reside in the United States, have been active on the Facebook platform at least once in the previous 30 days, have at least 100 U.S.- based Facebook friends, have a non-missing residential ZIP code, and for whom we are able to allocate at least one friend to a setting using the algorithm described in Section VI.B. The vertical axis shows the relative share of friends made in each of the six settings we analyze (e.g., high schools), defined as the average fraction of friends made in that setting by people in a given SES ventile (5 percentile rank bin) divided by the fraction of friends made in that setting in the whole sample. Numbers above 1 imply that people at a given SES rank make more friends in a given setting than the average person; numbers below 1 imply the opposite. Table 5 lists the underlying shares of friendships made in each setting for below-median-SES and above-median-SES people.

FIGURE 2: Economic Connectedness, Exposure, and Bias by Setting and Socioeconomic Status

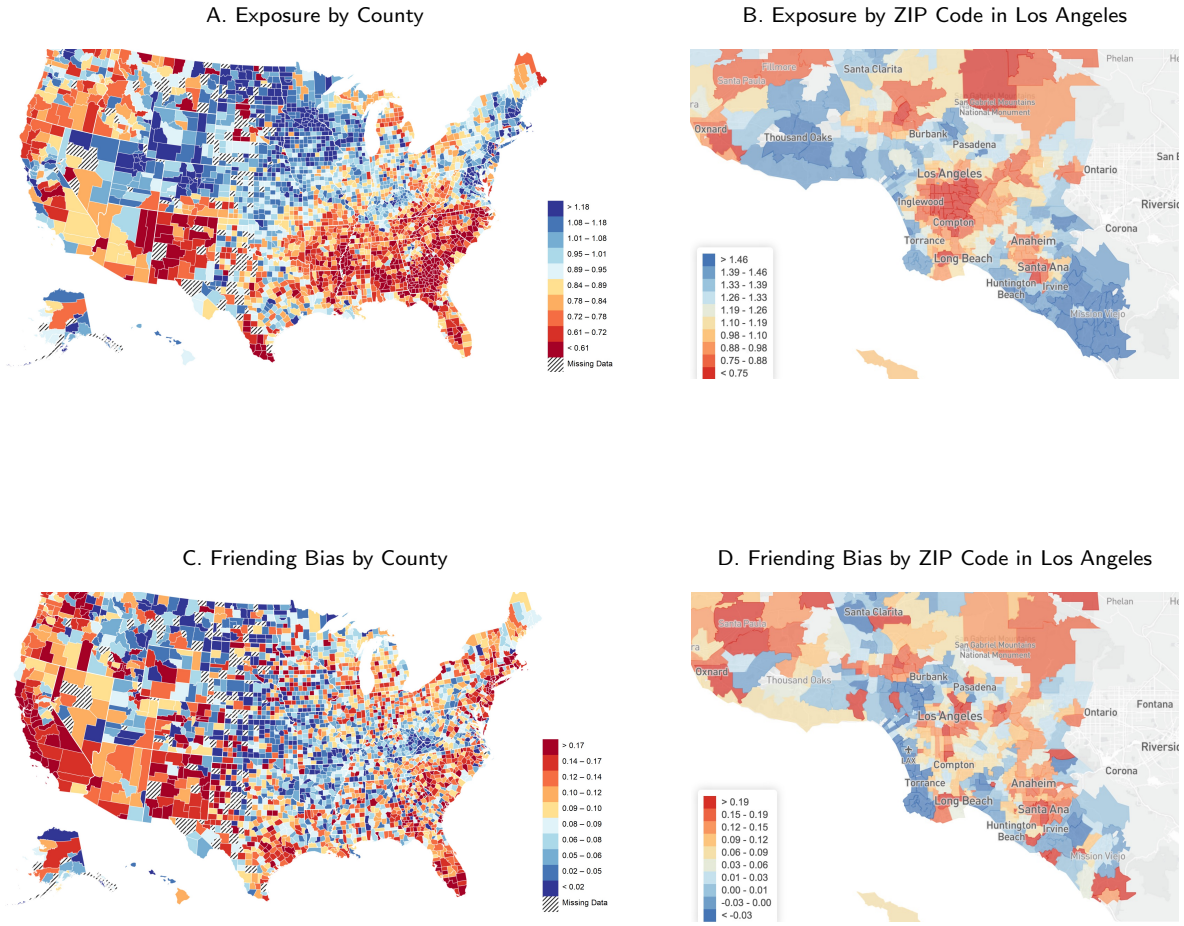


Notes: This figure shows how economic connectedness, exposure, and friending bias vary across the six settings we analyze by individuals' SES. All panels in this figure are based on the primary analysis sample defined in the notes to Figure 1. Panel A shows economic connectedness (EC) by setting for below-median-SES individuals (left, green bars) and above-median-SES individuals (right, orange bars). For both low- and high-SES individuals, EC is defined as twice the fraction of above-median-SES friends made within each setting. Panel B shows the mean rate of exposure to high-SES individuals in an individual's group (e.g., their high school) by setting for below-median-SES individuals (left, green bars) and above-median-SES individuals (right, orange bars). High-SES exposure is defined as two times the fraction of above-median-SES members of the individual's group. Panel C shows mean friending bias by setting for below-median-SES individuals (left, green bars) and above-median-SES individuals (right, orange bars). Friending bias is defined as one minus the ratio of the share of above-median-SES friends to the share of above-median-SES peers in the individual's group. EC, high-SES exposure, and friending bias are all calculated at the individual level and then aggregated to the setting x SES level (see Supplementary Information B.5 for details). Panel D restricts the sample to low-SES members of religious groups and plots these individuals' friending bias in each of the other settings minus their friending bias in religious groups. Table 5 lists the values of average economic connectedness, bias, and exposure shown in this figure.

FIGURE 3: Determinants of Differences in Economic Connectedness by SES and Across ZIP Codes

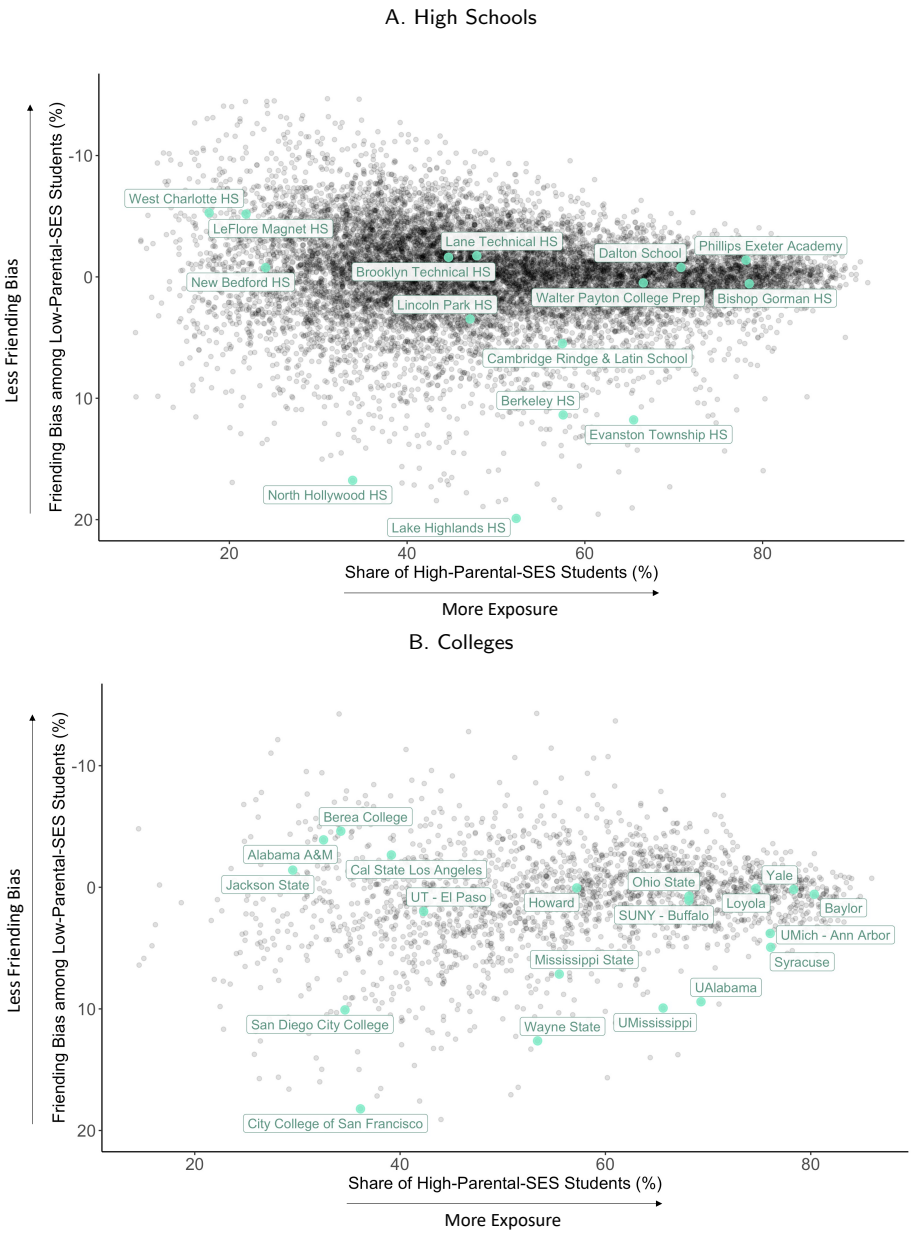
Notes: Panel A of this figure shows how much of the difference in the economic connectedness (EC) between high- vs. low-SES individuals is driven by differences in the settings in which they make friendships (friending rates), rates of exposure to high-SES individuals in those settings, and friending bias conditional on exposure. The first and fifth bars show the observed EC for average low- and high-SES individuals, calculated as the EC for individuals who have setting-level friending rates, exposure rates, and friending bias levels that match the means for low- and high-SES people in our sample, respectively (see Section VI.D). The middle three bars show the predicted economic connectedness for the average low-SES individual under various counterfactual scenarios. In the second bar, we consider a counterfactual scenario in which the friending rates across different settings for the average low-SES individual are equated to those of the average high-SES individuals, while preserving exposure and friending bias at the mean observed levels for low-SES individuals within those settings. The third bar further equates the rate of high-SES exposure in each setting to match the observed mean values for high-SES individuals. The fourth bar equates rates of friending bias in each setting as well as friending rates across settings to match the observed mean values for high-SES individuals. The fifth bar equates rates of both exposure and friending bias within settings and friending rates across settings. Panel B presents a decomposition exercise analogous to Panel A between ZIP codes with different levels of EC for below-median-SES residents instead of between below-median-SES and above-median-SES individuals. The comparison of interest here is between ZIP codes in the bottom quintile of the EC distribution for below-median-SES residents (Low-EC ZIP codes) and ZIP codes in the top quintile of EC for below-median-SES residents (High-EC ZIP codes). See Supplementary Information B.5 for further details on these counterfactual exercises.

FIGURE 4: The Geography of Exposure and Friending Bias



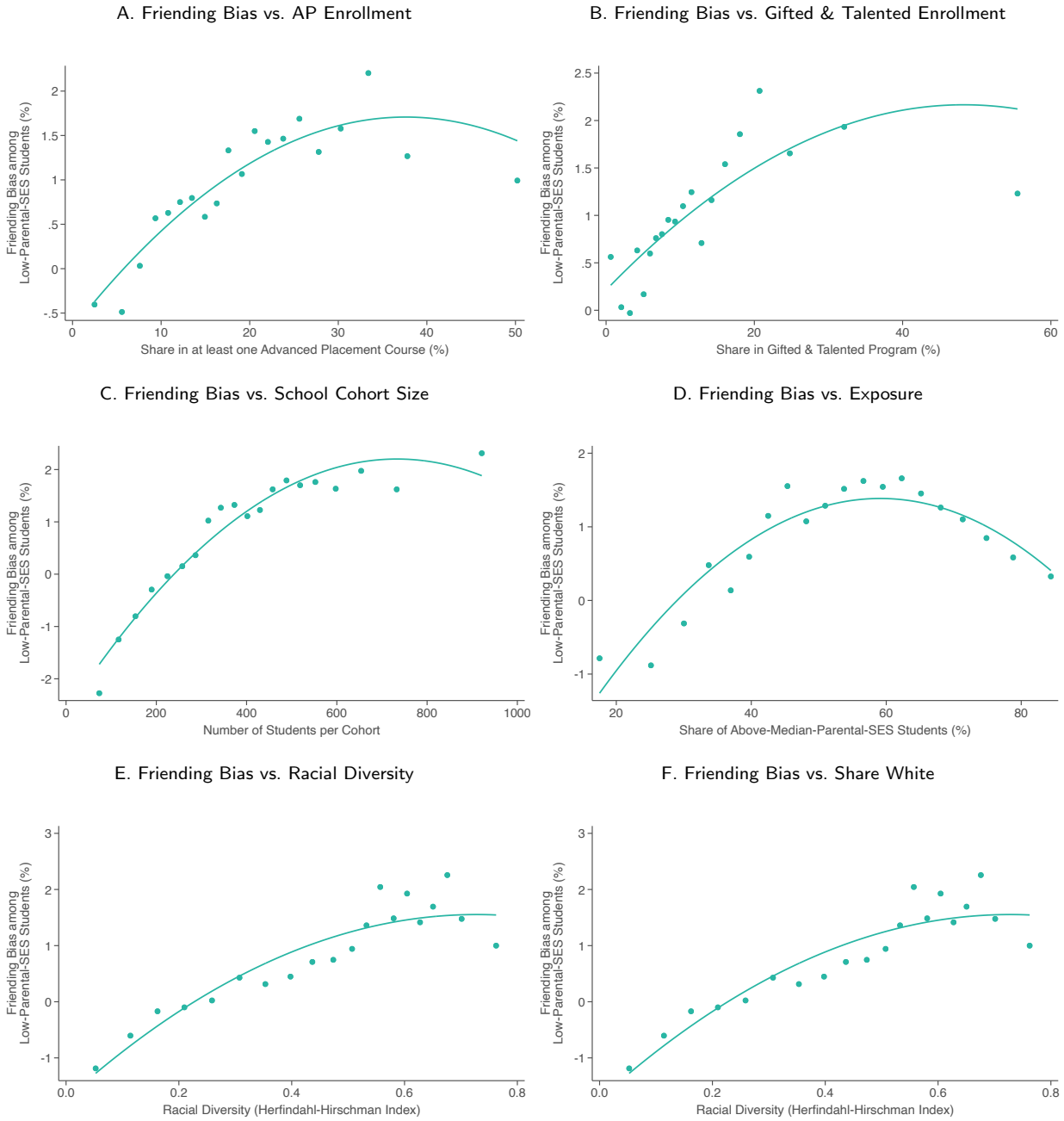
Notes: This figure shows maps of mean high-SES exposure (Panels A and B) and mean friending bias (Panels C and D) for low-SES individuals. These maps must be viewed in color to be interpretable. Panels A and C are national county-level maps; Panels B and D are ZIP code-level maps of the Los Angeles metropolitan area. We aggregate individual-level statistics to compute ZIP code-level and county-level means (see Supplementary Information B.5). At the individual level, exposure is defined as the weighted average of two times the fraction of individuals with above-median SES in the groups in which a below-median-SES individual participates, weighting each group by the individual's share of friends in that group. Friending bias is defined as one minus the weighted average of the ratio of the share of high-SES friends to the share of high-SES peers in the groups in which a low-SES individual participates, again weighting each group by the individual's share of friends in that group. We use methods from the differential privacy literature to add noise to the statistics plotted here to protect privacy while maintaining a high level of statistical reliability; see www.socialcapital.org for further details on these procedures.

FIGURE 5: Friending Bias and Exposure by High School and College



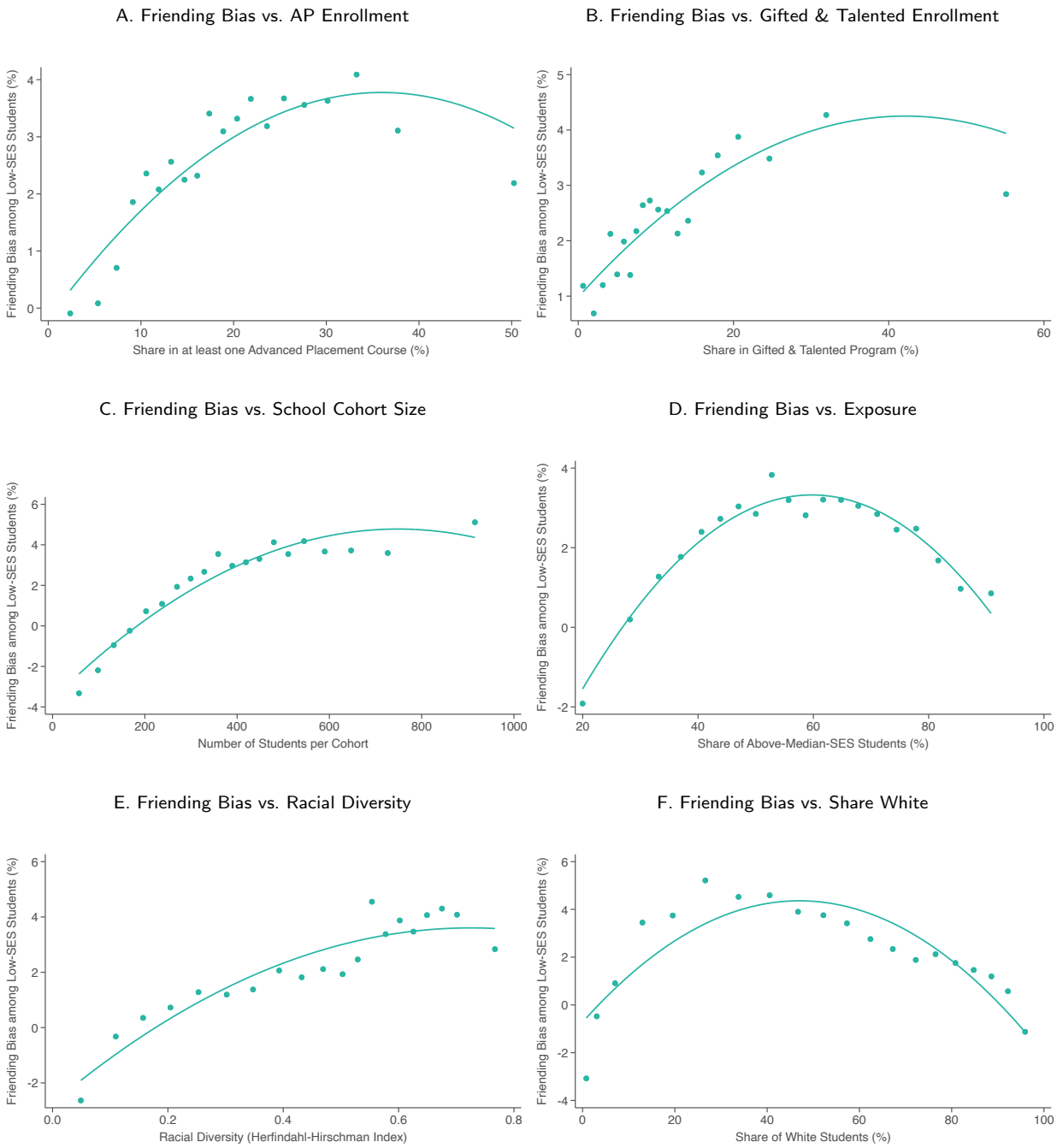
Notes: This figure plots mean friending bias among low-parental-SES students vs. the share of high-parental-SES students, by high school (Panel A) and college (Panel B). Friending bias is defined as one minus the mean ratio of the share of high-parental-SES high school friends to the share of high-parental-SES high school peers, averaging over low-parental-SES students (see Supplementary Information B.5). The vertical axis is reversed, so that schools and colleges in the upper half of each panel have lower friending bias. The sample consists of individuals in the 1990-2000 birth cohorts (approximately spanning the high school and college graduating classes of 2008-18 and 2012-22, respectively) who could be linked to a specific school or college and to parents with an SES prediction. We only report statistics for high schools and colleges that have at least 100 low-SES and 100 high-SES Facebook users summing across these cohorts. We use methods from the differential privacy literature to add noise to the statistics plotted here to protect privacy while maintaining a high level of statistical reliability; see www.socialcapital.org for further details on these procedures. In this figure, SES refers to the SES of individuals' parents; Supplementary Figure 2 replicates these figures using individuals' own (post-high school and post-college) SES ranks in adulthood.

FIGURE 6: Predictors of Friending Bias in High Schools Using Parental SES



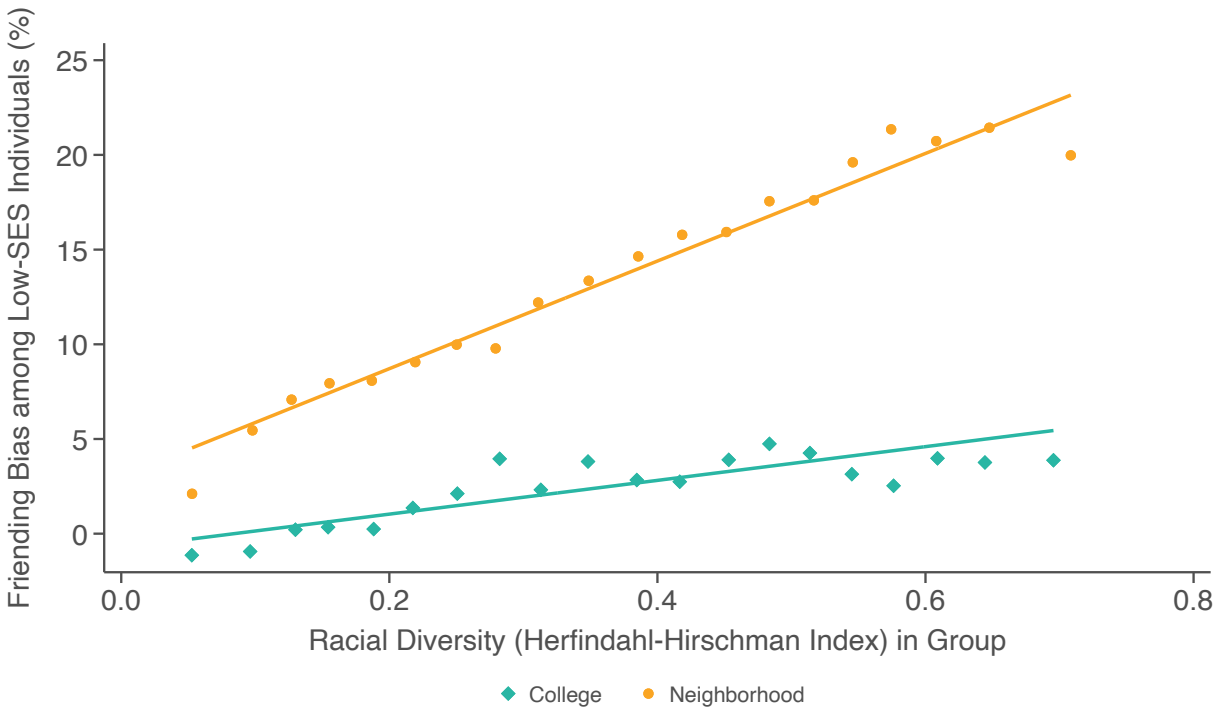
Notes: This figure shows school-level binned scatter plots of the average degree of friending bias (based on parental SES) among low-SES individuals vs. various school-level characteristics: the share of students enrolled in at least one Advanced Placement course (Panel A); the share of students enrolled in a Gifted and Talented program (Panel B); total number of students per cohort (Panel C); the share of above-median-parental-SES students (Panel D); an index of racial diversity defined as $100 \times (1 - \Sigma_i s_i^2)$, where s_i is the fraction of race/ethnicity i (Black, white, Asian, Hispanic, Native American) in the school (Panel E); and the share of white students (Panel F). Friending bias is defined as one minus the mean ratio of the share of high-SES high school friends to the share of high-SES high school peers, averaging over low-SES students in the 1990-2000 birth cohorts (see Supplementary Information B.5). Shares of students enrolled in at least one AP course and in a Gifted and Talented program are obtained from the 2015-2016 Civil Rights Data Collection (CRDC). School size and racial composition data are from the 2017-2018 National Center Education Statistics (NCES) data (see Supplementary Information A.2). The set of schools used in these plots is the same as in Figure 5, conditional on being present in the CRDC data. To construct the binned scatter plots, we divide the variable on the horizontal axis into ventiles (5 percentile point bins) and plot the mean of the vertical-axis variable vs. the mean of the horizontal-axis variable in each ventile. All binned scatter plots are weighted by the number of students in each high school as reported in the NCES data. As a visual guide to approximate the non-parametric relationships, the solid lines in each figure show lines of best fit from quadratic regressions estimated using OLS.

FIGURE 7: Predictors of Friending Bias in High Schools Using Own SES



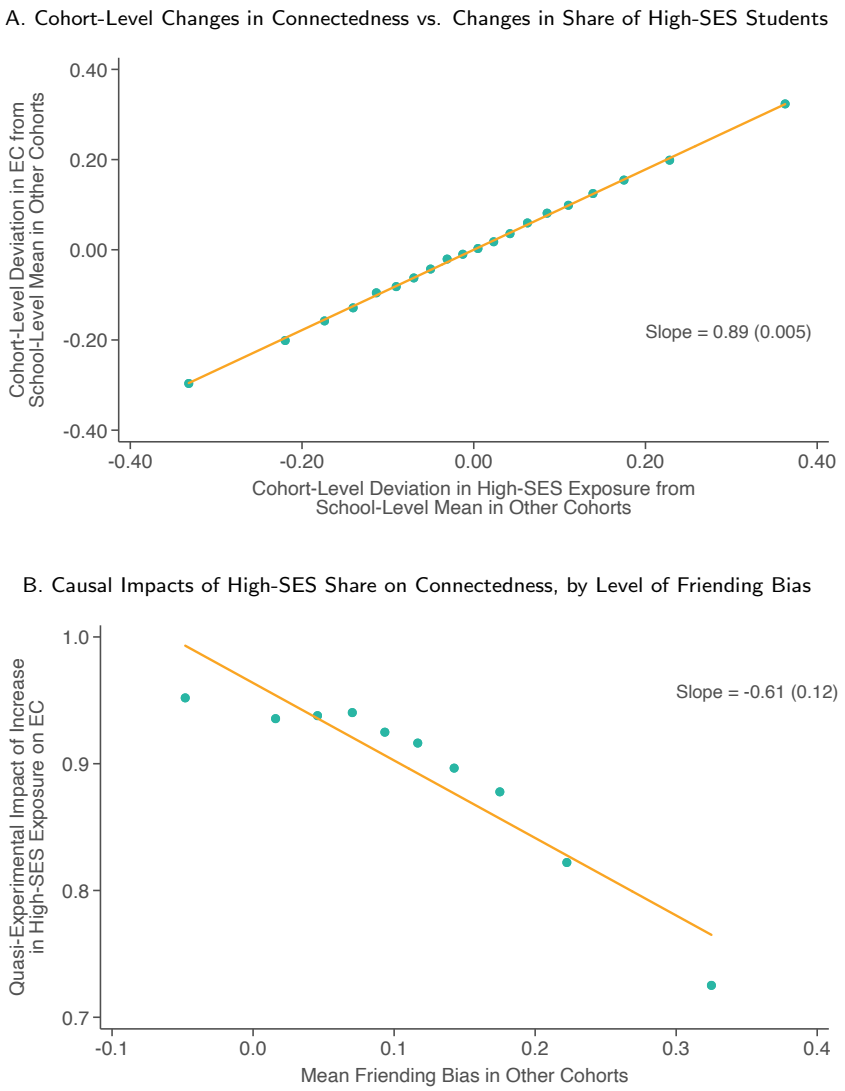
Notes: This figure replicates Figure 6 using own SES rank in adulthood instead of parental SES rank when measuring friending bias and when defining the share of above-median SES students in Panel D. See notes to Figure 6 for further details.

FIGURE 8: Friending Bias vs. Racial Diversity



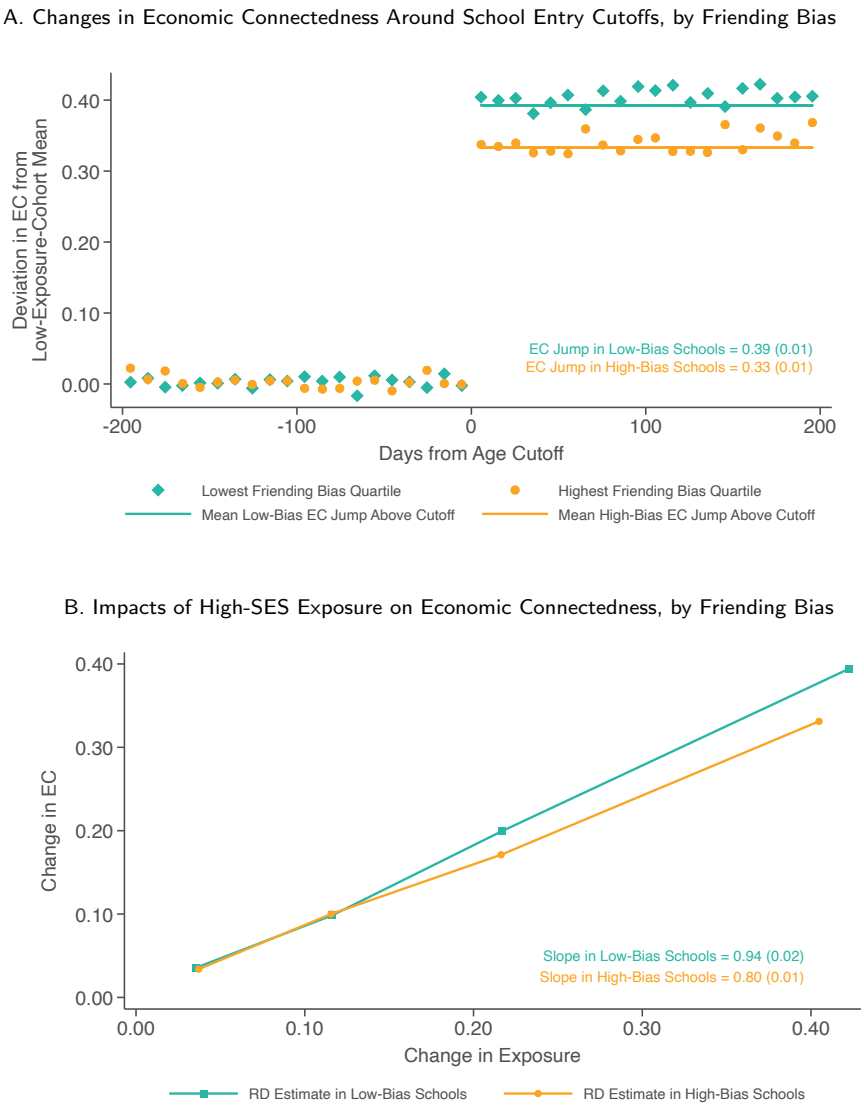
Notes: This figure presents binned scatter plots of friending bias vs. racial diversity within colleges (green diamonds) and neighborhoods (ZIP codes, orange circles). See notes to Figure6 for details on construction of binned scatter plots. We define racial diversity as $100 \times (1 - \Sigma_i s_i^2)$, where s_i is the fraction of race i (Black, white, Asian, Hispanic, Native American). Friending bias is defined as the average among below-median-SES individuals in the group (i.e., college or neighborhood) of one minus the ratio of the share of above-median-SES friends in each individual's group to the share of above-median-SES peers in that group. For comparability, both series measure SES using own SES rank in adulthood and use data from the 1986-1996 birth cohorts. Racial shares for each college and ZIP code are obtained from the 2013 Integrated Post-Secondary Education Data System (IPEDS) and the 2018 American Community Survey (ACS), respectively.

FIGURE 9: Causal Effects of Socioeconomic Integration on Economic Connectedness in High Schools: Cross-Cohort Estimates



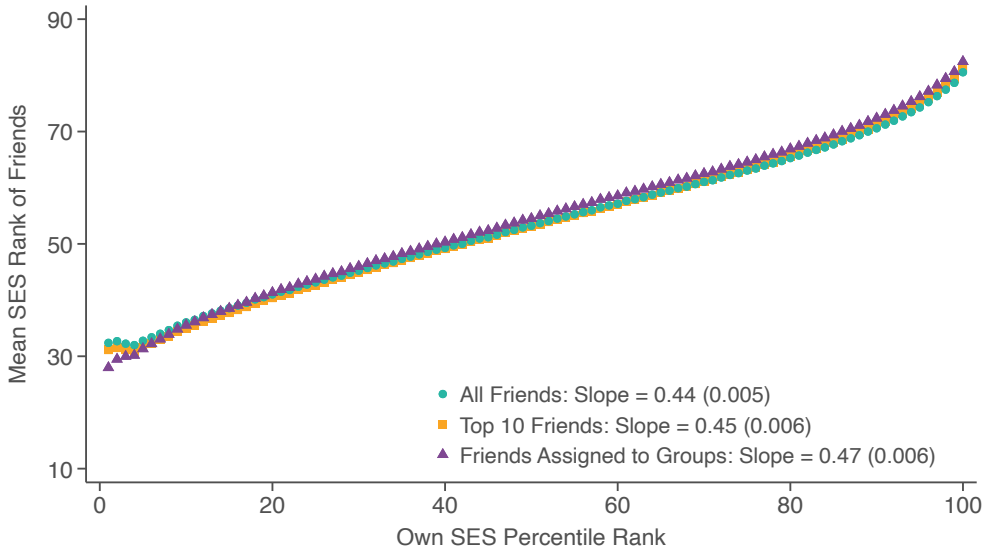
Notes: This figure analyzes the causal effect of being assigned to a high school cohort with more high-SES peers on the economic connectedness of low-SES students, based on the level of friending bias in the school (see Section VI.H for details). We measure economic connectedness (EC), exposure, and bias in this figure based on parental SES. The sample consists of all individuals in our primary analysis sample born between 1990 and 2000 whom we can link to parents and match to high schools. We further limit the sample to schools with at least 500 students (pooling all cohorts), at least 100 bottom-quintile-SES students, and at least 100 top-quintile-SES students. For each cohort, exposure is defined as five times the fraction of top-quintile-SES students; EC is defined as five times the average share of top-quintile-SES friends among bottom-quintile-SES students; and friending bias is defined as the average among bottom-quintile-SES students of one minus the ratio of the share of top-quintile-SES friends to the share of top-quintile-SES peers in their cohort. Panel A presents a binned scatter plot of the cohort-level deviations from school means in EC vs. cohort-level deviations from school means in exposure. The cohort-level deviations are constructed as the mean for the relevant cohort c in a given school minus the mean for all other cohorts in the same school, weighting by the number of bottom-quintile-SES students in each cohort. The binned scatter plot is constructed by dividing the cohort-level deviations in exposure into twenty equal-sized bins and plotting the mean deviation in EC vs. the mean deviation in exposure within each bin. We also report a slope estimated using a linear regression, with standard error clustered by high school in parentheses. To construct Panel B, we first divide school x cohort cells into ten deciles based on the mean level of friending bias for all other cohorts in the same school. We then estimate regressions analogous to that in Panel A using the school x cohort cells in each of the ten deciles separately. Finally, we plot the slopes from the ten regressions vs. the mean level of friending bias (leaving out the focal cohort) in each decile.

FIGURE 10: Causal Effects of Socioeconomic Integration on Economic Connectedness in High Schools: Regression-Discontinuity Estimates



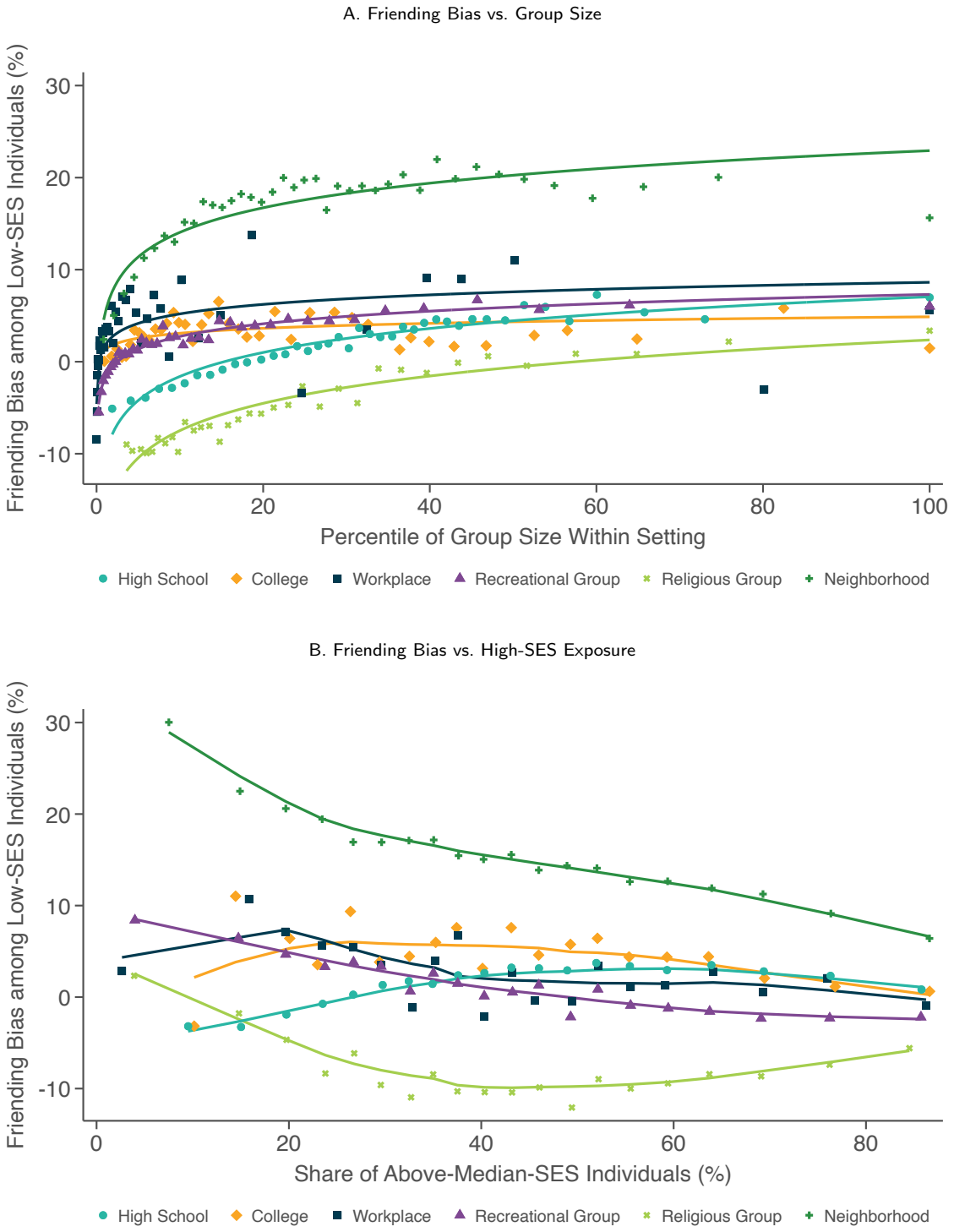
Notes: This figure analyzes the causal effects of being assigned to a high school cohort with more high-SES peers on economic connectedness (EC) using a regression discontinuity design, separately for schools with low vs. high levels of friending bias (see Section VI.I for details). The sample consists of all individuals in our analysis sample born between 1990 and 2000 whom we can link to their parents and match to high schools. Panel A shows the reduced-form impacts of jumps in the share of high-SES students on EC, separately by the level of friending bias. To construct Panel A, we first calculate the absolute values of changes in the fraction of students with top-quintile-parental-SES ("high-SES") across all consecutive high-school cohort pairs. We then restrict the sample to cohort pairs in the top quartile of this distribution of exposure changes, and order cohorts so that the cohort with the lower share of high-SES students is the first of the two cohorts. We define EC for a given student as five times the share of top-quintile-parental-SES friends in their cohort. We then calculate means of EC for all bottom-quintile-parental-SES students in a cohort pair, pooling students by their age distance in days from the cohort age cutoff (subtracting the mean EC among low-SES students in the cohort). We do this separately for schools in the top and bottom quartile of the distribution of friending bias, calculated as the average friending bias of low-SES individuals in the same high school over all cohorts excluding the own and adjacent cohorts. We report regression discontinuity (RD) estimates (with standard errors in parentheses) of the jump in average EC at the cutoff estimated using a linear regression with a bandwidth of 200 days; the solid lines plot the magnitudes of these jumps (see Section VI.I). In Panel B, we plot RD estimates of the jump in EC (estimated as in Panel A) for each of the four quartiles of changes in exposure across cohorts against the mean change in (normalized) high-SES exposure in that exposure-jump quartile. We again plot these estimates separately for schools in the top and bottom quartile of leave-out friending bias.

FIGURE 11: Relationship Between Friends' Socioeconomic Status and Own Socioeconomic Status



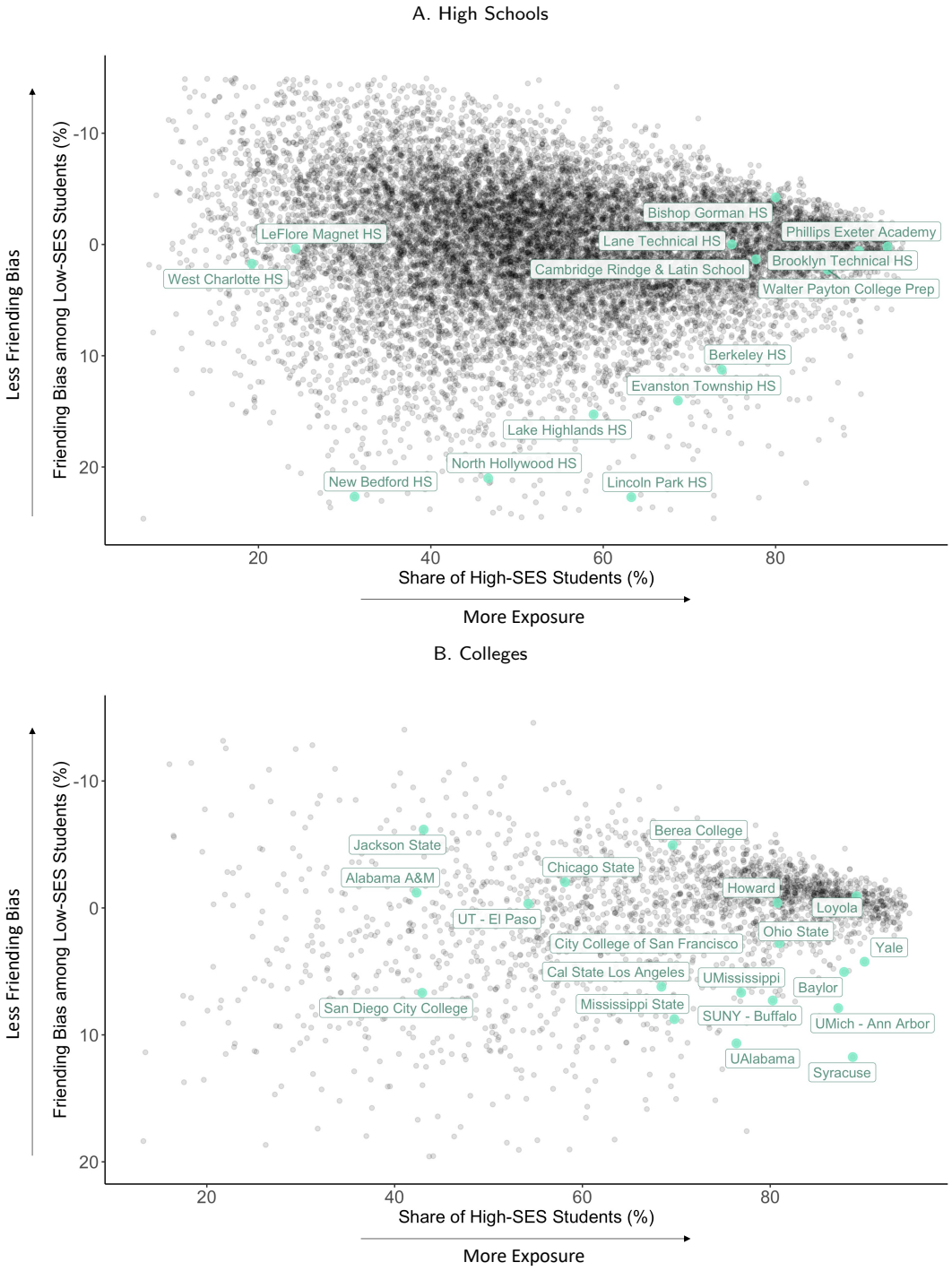
Notes: This figure plots the mean socioeconomic status (SES) percentile rank of individuals' friends vs. their own SES percentile rank. The series in green circles is calculated using the entire friendship network for each individual. The series in orange squares is constructed using each individual's ten closest friends, based on the frequency of public interactions such as likes, tags, wall posts, and comments. The green and orange series are identical to those in Figure 1 of Chetty et al. 2022. The series in purple triangles replicates the series in green circles using the 30% of friendships that we are able to assign to a group, the primary analysis sample used in this paper. For each series, we report slopes estimated from a linear regression on the plotted points, with heteroskedasticity-robust standard errors in parentheses.

SUPPLEMENTARY FIGURE 1: Predictors of Friending Bias across Settings



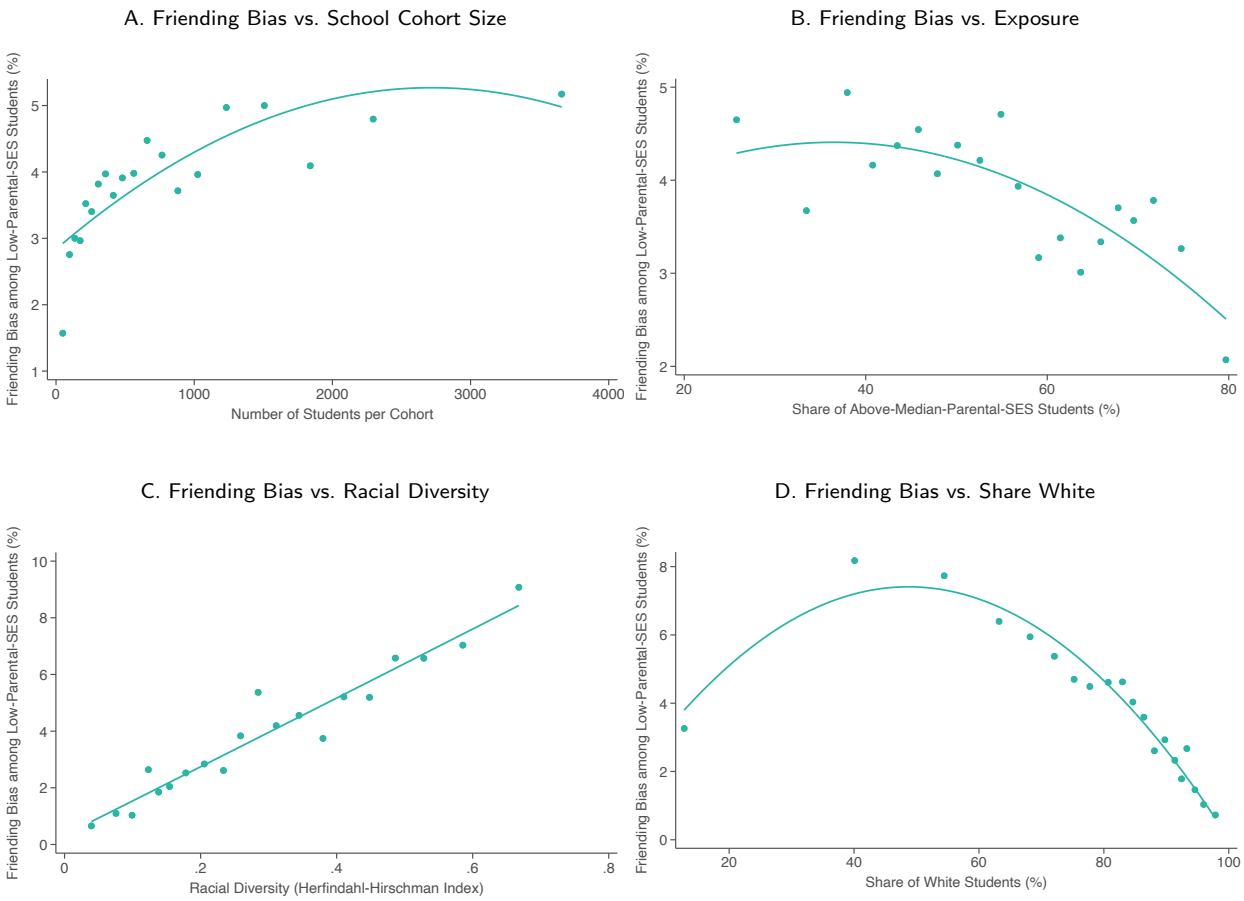
Notes: Panel A plots mean friending bias among below-median-SES individuals vs. group size, measured as a percentile in the within-setting group size distribution, weighting by the number of low-SES individuals. Panel B presents a binned scatter plot of friending bias vs. the share of above-median SES individuals in the group, again weighting by the number of low-SES individuals. Both friending bias and the share of high-SES individuals are computed using the 1986-1996 birth cohorts and using individuals' own SES. Unlike Figures 6 and 7 and Supplementary Figure 3, we do not impose any restrictions on the set of high schools and colleges included in this figure.

SUPPLEMENTARY FIGURE 2: Friending Bias and Exposure by High School and College, Based on Own SES



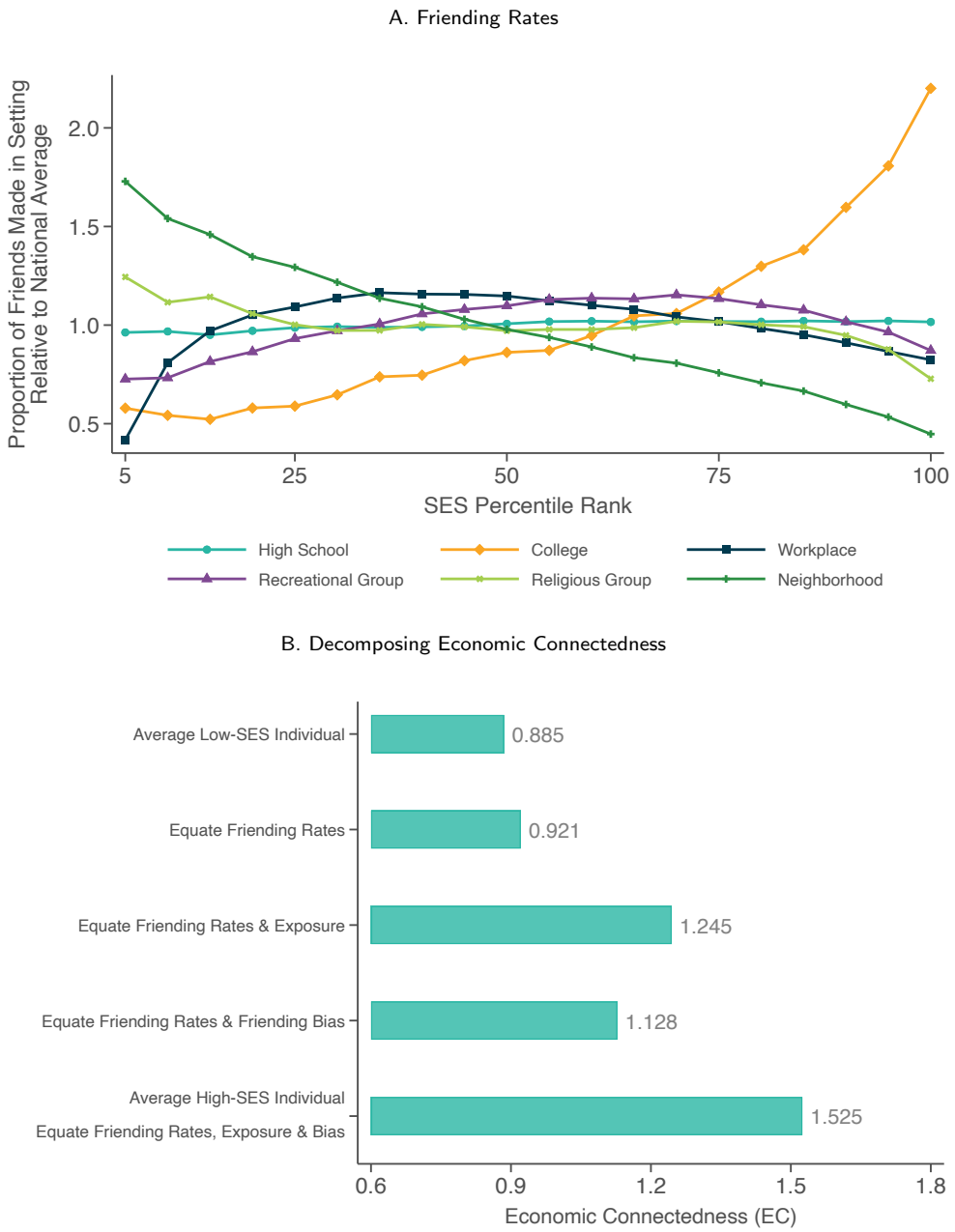
Notes: This figure replicates Figure5 using own (post-high-school) SES rank instead of parental SES rank. In this figure, we focus on the 1986-1996 birth cohorts (aged between 26–36 in 2022) to obtain more reliable measures of SES ranks in adulthood. We label the same schools that are labeled in Figure 5, unless they fail to meet the minimum threshold of 100 low-SES and 100 high-SES Facebook users when using individuals' own SES ranks. For example, some elite private high schools meet this threshold for parental SES but not own SES, as their students' own SES in adulthood tends to be significantly higher than that of their parents.

SUPPLEMENTARY FIGURE 3: Predictors of Friending Bias in Colleges



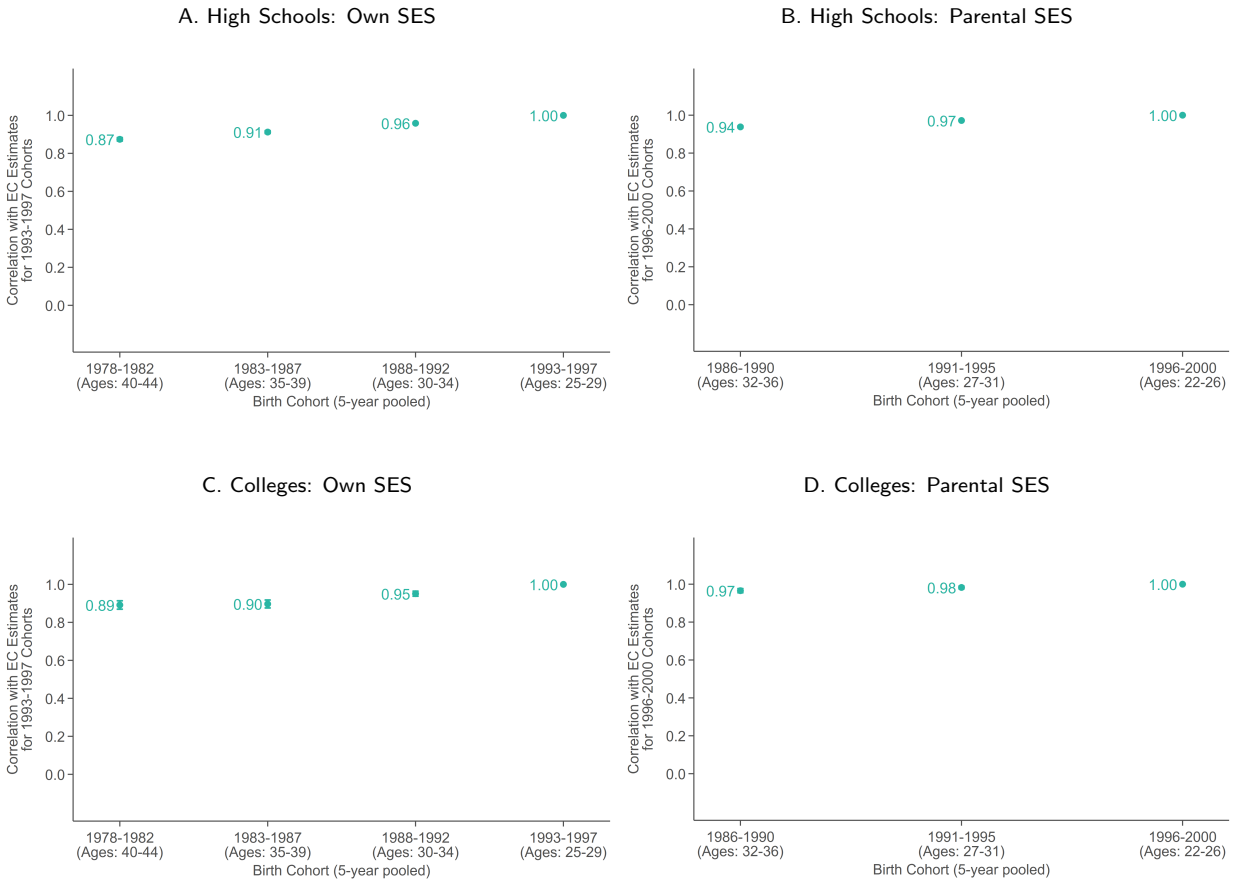
Notes: This figure replicates Figures 6c-f in colleges instead of high schools. See notes to Figure 6 for details.

SUPPLEMENTARY FIGURE 4: Friending Rates and Determinants of Economic Connectedness, Correcting for Underreporting of Group Memberships



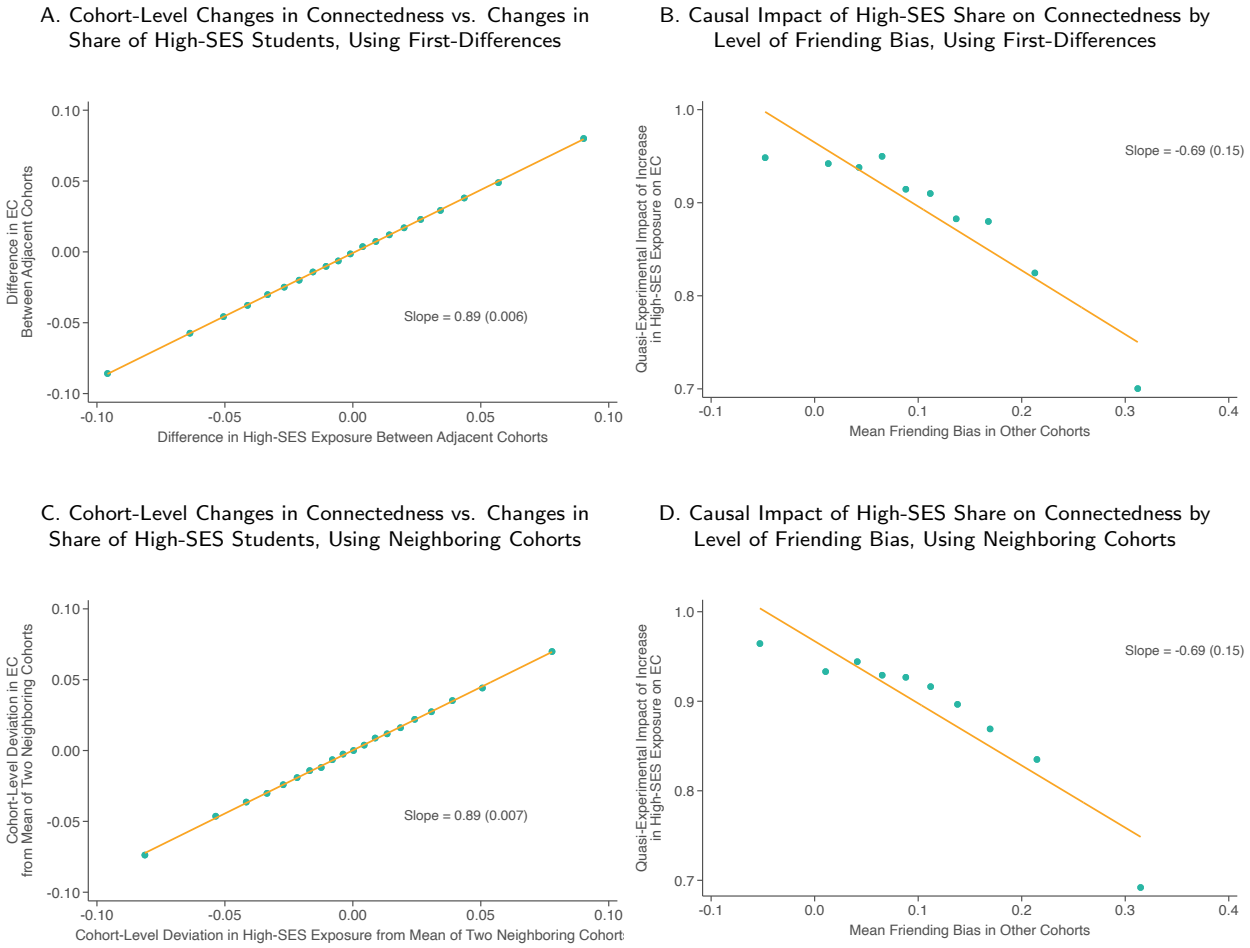
Notes: Panel A replicates Figure 1 and Panel B replicates Figure 3a correcting for underreporting of group memberships by Facebook users using the algorithm described in Supplementary Information B.3, which inflates the raw share of friends made across groups by group membership rates estimated from external data.

SUPPLEMENTARY FIGURE 5: Autocorrelation of EC by Birth Cohort Across High Schools and Colleges



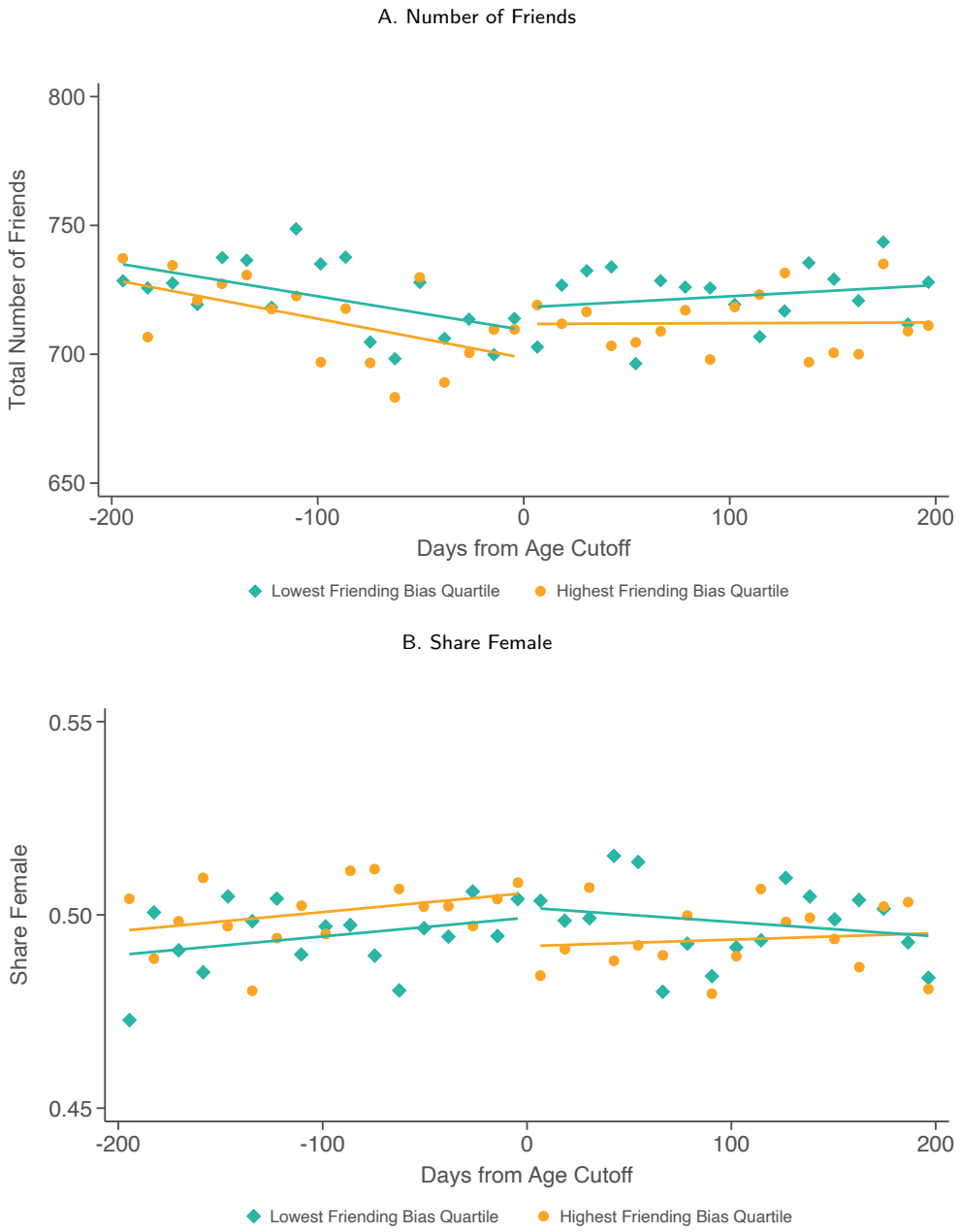
Notes: Panel A plots the serial correlation across high schools of EC estimated using individuals' own SES different birth cohorts. Each point shows the correlation between estimates of EC based on a given 5 year birth cohort range, starting from 1978-1982, with the reference cohort group of 1993-1997. We report the ages when SES is measured (in 2022) below each cohort group. Panel B replicates Panel A using parental SES, with 5-year intervals starting from the 1986-1990 cohorts and a reference cohort group of 1996-2000. Panels C and D replicate Panels A and B at the college level. All correlations are weighted by the average number of low-SES students in the relevant pair of five-year cohort groups. When analyzing own SES (Panels A and C), we restrict the sample to schools that meet the following size restrictions over the relevant 5-year intervals: more than 75 low-SES students, more than 75 high-SES students, and more than 200 total students. When analyzing parental SES (Panels B and D), we require more than 10 low-SES students, more than 10 high-SES students, and more than 25 total students.

SUPPLEMENTARY FIGURE 6: Causal Effects of Changes in Socioeconomic Integration on Economic Connectedness in High Schools: Sensitivity Analysis



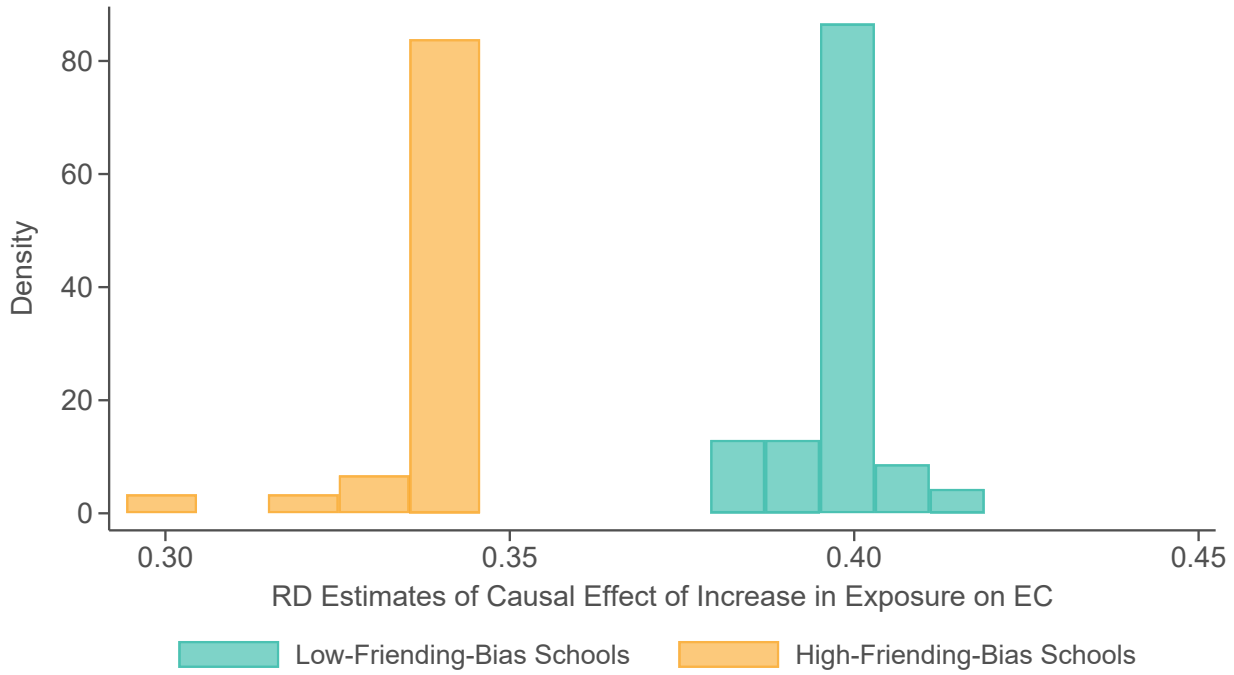
Notes: This figure replicates Figure 9 using alternative measures of changes in exposure and economic connectedness across cohorts. Panels A and B use differences in exposure and economic connectedness between adjacent cohorts, measuring friending bias in cohorts excluding the two adjacent cohorts used to compute changes in exposure and EC. Panels C and D demean exposure and EC using the mean of the two neighboring cohorts, instead of the mean over all other cohorts in the school as in Figure 9. See notes to Figure 9 for further details.

SUPPLEMENTARY FIGURE 7: Balance Tests for Regression Discontinuity Design



Notes: This figure evaluates the identification assumption underlying the regression discontinuity design by replicating Figure 10a using other variables: the total number of friends individuals have in the primary analysis sample (Panel A) and an indicator for being female (Panel B). See notes to Figure 10 for details on the construction of this figure.

SUPPLEMENTARY FIGURE 8: RD Estimates of Causal Effects of Exposure on EC, by Bandwidth



Notes: Figure 10a reports regression discontinuity estimates of the causal effect of changes in exposure on EC separately for low-bias and high-bias schools using linear regressions with a bandwidth of 200 days around the school entry cutoff (see Supplementary Information B.6 for the regression specification). This figure presents histograms of estimates from the same regression specification, varying the bandwidth around the school entry cutoff from 20 to 300 days in 10 day increments, separately for low-bias and high-bias schools.