

NBER WORKING PAPER SERIES

SOCIAL CAPITAL I:
MEASUREMENT AND ASSOCIATIONS WITH ECONOMIC MOBILITY

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 30313
<http://www.nber.org/papers/w30313>

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)). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by 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. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission

Social Capital I: Measurement and Associations with Economic Mobility

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. 30313

July 2022

JEL No. R0

ABSTRACT

In this paper—the first in a series of two papers that use data on 21 billion friendships from Facebook to study social capital—we measure and analyze three types of social capital by ZIP code in the United States: (i) connectedness between different types of people, such as those with low vs. high socioeconomic status (SES); (ii) social cohesion, such as the extent of cliques in friendship networks; and (iii) civic engagement, such as rates of volunteering. These measures vary substantially across areas, but are not highly correlated with each other. We demonstrate the importance of distinguishing these forms of social capital by analyzing their associations with economic mobility across areas. The fraction of high-SES friends among low-SES individuals—which we term economic connectedness—is among the strongest predictors of upward income mobility identified to date, whereas other social capital measures are not strongly associated with economic mobility. If children with low-SES parents were to grow up in counties with economic connectedness comparable to that of the average child with high-SES parents, their incomes in adulthood would increase by 20% on average. Differences in economic connectedness can explain well-known relationships between upward income mobility and racial segregation, poverty rates, and inequality. To support further research and policy interventions, we publicly release privacy-protected statistics on social capital by ZIP code 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

Social capital—the strength of an individual’s social network and community—has been identified as a central determinant of outcomes in many fields, from economics to education to health (e.g., Eagle et al. 2010, Carrell et al. 2011, Sacerdote 2011, Beaman 2012, Laschever 2013, Aral and Nicolaides 2017, Hedefalk and Dribe 2020, List et al. 2020), and recent work has argued that social capital may shape important social phenomena, such as income inequality and economic opportunity (e.g., Putnam 2016, Social Capital Project 2019). However, a lack of large-scale data on social networks has limited researchers’ ability to understand what types of social capital matter for such outcomes and how we can increase impactful forms of social capital. Most existing studies of social capital rely either on relatively small surveys or on datasets that only cover specific communities (e.g., a single college), limiting researchers’ ability to learn about the determinants and effects of social capital by comparing different social settings.¹

We use data on 21 billion friendships from Facebook to construct and analyze measures of social capital for each ZIP code, high school, and college in the United States.² Building on surveys by DeFilippis (2001), Jackson (2020), and Portes (1998), we organize measures of social capital proposed in prior work into three categories:

1. Cross-type *connectedness*: the extent to which different types of people (e.g., high- vs. low-income) are friends with each other (e.g., Loury 1977, Bourdieu 1986, Lin and Dumin 1986, Putnam 2016);
2. Network *cohesiveness*: the degree to which friendship networks are clustered into cliques and whether friendships are supported by mutual friends (e.g., Coleman 1988);
3. *Civic engagement*: indices of trust or participation in civic organizations (e.g., Putnam 1995; Putnam et al. 1994).

¹For example, the most widely used dataset to study social networks—the National Longitudinal Study of Adolescent to Adult Health (Add Health)—covers approximately 20,000 students at 132 schools and cannot be disaggregated by school due to small sample sizes. A more nascent strand of literature uses large scale mobile phone data to measure “experienced segregation” (e.g., Eagle et al. 2010; Wang et al. 2018; Reme et al. 2022; Athey et al. 2020; Dong et al. 2020; Levy et al. 2020), but does not directly measure social interactions between different types of people as we do here, a distinction that proves to be empirically important.

²Facebook data have been used to study the effects of social networks on a variety of outcomes in prior work: patent citations (Bailey et al. 2018a), home purchasing decisions (Bailey et al. 2018b), mortgage choices (Bailey et al. 2019a), cell phone adoption (Bailey et al. 2019b), labor market outcomes (Gee 2018; Gee et al. 2017), commuting flows (Bailey et al. 2020a), international trade flows (Bailey et al. 2021), investment decisions (Kuchler et al. 2022a), peer-to-peer lending (Allen et al. 2020), EITC claiming behavior (Wilson 2020), racial homophily (Wimmer and Lewis 2010), health behavior and beliefs (Bailey et al. 2020b), mortality rates (Hobbs et al. 2016), the spread of COVID-19 (Kuchler et al. 2022b), and the social integration of international migrants (Bailey et al. 2022). Although they use some of the same underlying data, prior studies have not constructed systematic measures of social capital or studied their determinants as we do here. As in prior work, we use social network data to proxy for “real world” friendships rather than online interactions per se; as a result, our analysis does not shed light on the effects of online social networks themselves.

In addition to measuring distinct concepts, these categories of social capital differ in terms of the data they use as inputs: the cross-type connectedness measures combine data on networks (friendship links) with data on individual characteristics; the cohesiveness measures only use data on the network links, with no characteristics; and the civic engagement measures do not use data on networks at all and are instead based purely on individual or community-level characteristics.

In the first part of this paper, we construct several measures of social capital at the national, county, and ZIP code levels using data on the social networks of 72.2 million Facebook users between the ages of 25–44. We publicly release and visualize privacy-protected statistics on all the social capital measures at www.socialcapital.org.³

We begin by measuring connectedness across different types of people, focusing on a new measure that we term *economic connectedness*: the extent to which low- and high-socioeconomic status (SES) individuals are friends with each other.⁴ Consistent with prior work documenting homophily in friendships, we find that higher-SES individuals tend to have higher-SES friends: a 10 percentile point increase in one’s own SES rank in the national distribution is associated with a 4.4 percentile point increase in the mean SES rank of one’s friends. For example, individuals at the 20th percentile of the national SES distribution have friends with an average SES rank at the 40.9th percentile, while individuals at the 30th percentile have friends with an average SES rank at the 45.3th percentile. There is substantial variation in the degree of economic connectedness (EC) across areas. For example, low-SES individuals who live in the Midwest tend to have a higher share of high-SES friends than those who live in the Southeast. Furthermore, there is substantial local variation in EC even across ZIP codes within a given county. In addition to economic connectedness, we also measure connectedness across other characteristics, such as age and language.

We also measure network cohesiveness for each geography using concepts introduced in the prior literature on networks, such as clustering and support (measures of cliquishness) and spectral homophily (a measure of network fragmentation). We construct measures of civic engagement using Facebook data on group memberships, such as the rate at which individuals participate in volunteering groups and the density of civic organizations in an area. These statistics provide more precise and granular measures of the widely used proxies for civic engagement introduced by Putnam (1995) and further developed in Rupasingha et al. (2006).

³We also release measures of social capital by high school and college, as constructed in Chetty et al. (2022).

⁴We measure 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. We do not observe individuals’ incomes directly; however, we show that our SES measures are highly correlated with external, publicly available measures of average incomes across groups (e.g., ZIP codes, high schools, and colleges). We also show that simple measures of SES such as median household income in an individual’s ZIP code yield very similar results.

As with economic connectedness, all of these measures of social capital vary substantially across areas. However, the social capital measures are only weakly correlated with each other across space: areas with high levels of economic connectedness do not necessarily have highly cohesive networks or high civic engagement. The upshot is that one must specify the concept of “social capital” of interest to determine whether a given community is rich in social capital. This is particularly important because different types of social capital may be relevant for different outcomes.

In light of this result, in the second part of the paper, we investigate which notions of social capital are most associated with one outcome that has been the focus of recent academic research and public discussion: rates of upward economic mobility across generations. Using public statistics on economic mobility from Chetty et al. (2018), we find that economic connectedness is strongly associated with differences in upward income mobility across areas, with a correlation of 0.69 across ZIP codes. Network cohesiveness and civic engagement are not strongly associated with economic mobility on average: there are many communities that have tightly knit friendship networks or high levels of civic engagement, yet have low economic mobility. This is not because these other measures of social capital have no relationship with economic mobility, but because the relationships vary across areas. For example, in urban Chicago and Houston, ZIP codes with more cohesive networks (a greater fraction of friends who themselves are friends with each other) have significantly higher economic connectedness and mobility. In contrast, in suburban areas surrounding these cities, ZIP codes with more cohesive networks have lower economic connectedness and mobility. In multivariable regressions, EC remains highly predictive of upward mobility, while the other social capital measures add little explanatory power.

One reason that economic connectedness is associated with economic mobility may be that climbing out of poverty might be facilitated by connections to people who can shape aspirations or provide access to information and job opportunities, mechanisms proposed by Loury (1977) and Lin and Dumin (1986). This interpretation is consistent with Putnam’s (2000) argument that bridging capital—a concept that encompasses economic connectedness—is particularly valuable for “getting ahead.” Other forms of social capital may not provide such access outside one’s group and thus may matter less for economic advancement.

While this is one plausible explanation for the association between economic connectedness and mobility, there are also several alternative explanations for the correlation that do not rely on a causal effect of connectedness on mobility. In the third part of the paper, we analyze three such explanations, with the broader aim of better understanding the channels through which connectedness and mobility are related.

The first explanation we consider is reverse causality: greater economic mobility could lead to

greater economic connectedness because upwardly-mobile individuals who have high SES as adults may retain connections to lower-SES individuals they befriended as youth. To assess the importance of this channel, we construct measures of *childhood* economic connectedness, based on childhood (high school) friendships and parental SES, using data from Facebook as well as a supplementary sample from Instagram. Because friendships made in high school and parental SES are realized before individuals' incomes in adulthood, they cannot be directly influenced by rates of economic mobility. Childhood EC remains highly correlated with upward mobility, indicating that the correlation is not primarily driven by a causal effect of mobility on connectedness.

A second potential explanation for the correlation between connectedness and mobility is selection: the types of families who live in high-EC areas may inherently have higher rates of mobility. To assess the importance of such sorting, we first show that differences by race—perhaps the most salient form of residential sorting in the U.S.—do not explain our findings. EC remains highly correlated with race-specific measures of upward mobility for white, Black, and Hispanic individuals in racially homogeneous areas. We also evaluate the importance of sorting on other (potentially unobservable) dimensions using quasi-experimental estimates of the causal effect of growing up for an additional year in each county on earnings in adulthood constructed by Chetty and Hendren (2018b). Economic connectedness is strongly correlated with counties' *causal effects* on upward mobility, indicating that the correlation between EC and mobility is not simply driven by selection. The estimates imply that if low-SES children were to grow up in counties with economic connectedness comparable to the average high-SES child in the U.S., their incomes would increase by about 20% on average.

A third potential explanation for the correlation between economic connectedness and economic mobility is omitted variables: living in a place with higher economic connectedness may cause higher rates of mobility, but do so because of other area-level factors correlated with connectedness (e.g., better schools) rather than connectedness itself. As a step toward isolating the role of EC itself, we compare the relative explanatory power of EC and the strongest neighborhood-level predictors of economic mobility identified in prior work. Economic connectedness remains highly predictive of differences in economic mobility across areas even conditional on these other neighborhood characteristics, such as poverty rates, the quality of local schools, and family structure.⁵

Furthermore, the relationships between other neighborhood characteristics and mobility become much weaker once we control for EC, suggesting that the links between those factors and mobility may run through their impacts on economic connectedness. For example, controlling for connectedness

⁵While higher levels of EC are associated with better outcomes for children from low-SES families, they are uncorrelated with the outcomes of children from high-SES families once we control for local income levels. This finding raises the possibility that greater interaction across class lines can benefit the poor without offsetting negative effects for the rich.

greatly diminishes the association between poverty rates and mobility, suggesting that segregation by income may inhibit upward mobility insofar as it reduces contact with higher-SES people. Similarly, controlling for connectedness eliminates the negative correlation between racial segregation and economic mobility documented by Cutler and Glaeser (1997) as well as the negative correlation between income inequality and upward mobility (the “Great Gatsby Curve”) first documented by Corak (2013). While these findings do not definitively establish that EC has a causal effect on mobility—as there could be other factors correlated with both EC and mobility that have not yet been identified—they call for further analysis of its causal impacts, perhaps through interventions that directly increase connectedness. We turn to analyzing the determinants of economic connectedness and what form such interventions might take in the next paper in this series (Chetty et al. 2022).

While our analysis highlights connections to higher-SES people as a form of social capital particularly relevant to economic mobility, we caution that this finding does not imply that economic connectedness is the “best” or most important measure of social capital in general. Other forms of social capital may be more predictive of other outcomes, an issue that can be analyzed in future research using the measures made public here.

The paper is organized as follows. Section II presents our new measures of social capital. In Section III, we analyze the association between measures of social capital and economic mobility. Section IV analyzes the sources of the correlation between economic connectedness and mobility. 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.

II Measuring Social Capital

Building on prior work by DeFilippis (2001), Jackson (2020), and Portes (1998), we organize measures of social capital into three categories: (1) cross-type *connectedness*: the extent to which different types of people (e.g., high- vs. low-income) are friends with each other (Bourdieu 1986; Lin and Dumin 1986; Loury 1977; Putnam 2016); (2) network *cohesiveness*: the degree to which friendship networks are clustered into cliques and whether friendships tend to be supported by mutual friends (Coleman 1988); and (3) *civic engagement*: indices of trust or participation in civic organizations (Putnam 1995; Putnam et al. 1994).

Cross-type connectedness can be viewed as a form of “bridging” capital, while network cohesiveness is more in line with the concept of “bonding” capital (Putnam 2000). In addition to measuring distinct concepts, these categories of social capital differ in terms of the data they use as inputs: the cross-

type connectedness measures combine data on networks (friendship links) with data on individual characteristics; the cohesiveness measures only use data on the network links, with no characteristics; and the civic engagement measures do not use data on networks at all and are instead based purely on individual or community-level characteristics (Supplementary Table 1).

We measure these concepts, which we define more precisely below, using privacy-protected data from Facebook (see Sections VI.A and VI.H). 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 residential ZIP code. We focus on the 25–44 age range because its Facebook usage rate is above 80% (Perrin and Anderson 2019). Based on comparisons to nationally representative surveys and other supplementary analyses, we find that our Facebook analysis sample provides a reasonably representative picture of the national population (see Section VI.C). We use the Facebook data to obtain information on friendships, locations (ZIP code and county), and own and parental socioeconomic status; we describe these variables in detail in Section VI.B.

II.A Economic Connectedness

Many theoretical studies have shown how connections to more educated or affluent individuals can be valuable for transferring information, shaping aspirations, and providing mentorship or job referrals (Ambrus et al. 2014; Bolte et al. 2020; Bourdieu 1986; Calvo-Armengol and Jackson 2004; Jackson 2021; Lin 1999; Loury 1977; Montgomery 1991; Putnam 2016; Small 2010). Consistent with these models, empirical studies have documented that social ties to well-resourced individuals can materially impact economic and labor market outcomes (Beaman 2012; Burchardi and Hassan 2013; Laschever 2013; Sacerdote 2011). Motivated by this literature, we begin by measuring connectedness across different types of people, focusing on *economic connectedness*: the extent to which low- and high-socioeconomic status individuals are friends with each other.

Social scientists have measured “socioeconomic status” (SES) using many different variables, ranging from income and wealth to educational attainment, occupation, family background, neighborhood, and consumption (White 1982). To capture these varied definitions, we compute socioeconomic status for each individual in our analysis sample by combining several measures of SES, such as average incomes in the individual’s neighborhood and self-reported educational attainment (see Section VI.H for a discussion of how user privacy was protected during this project). We combine these measures of SES into a single SES index using a machine learning algorithm described in Section VI.B and further in Supplementary Information B.1. We then calculate each individual’s percentile rank in the national

SES distribution relative to others in their birth cohort. While we do not observe individuals’ incomes directly, we show that our SES rankings are highly correlated with external, publicly available measures of income across groups (e.g., ZIP codes, high schools, and colleges). We also show that using simpler measures of SES such as median household income in an individual’s ZIP code yields very similar results to those reported below.

Figure 1a plots the mean SES rank of individuals’ friends against their own SES ranks. There is strong homophily: higher-SES individuals have higher-SES friends. A one percentile point increase in one’s own SES rank is associated with a 0.44 percentile point increase in the SES rank of one’s friends on average. The relationship is almost perfectly linear between the 10th and 90th percentiles of the SES distribution, with a slope of 0.41 in that range. The slope rises to 0.98 between the 90th and 100th percentiles, showing that the highest-SES individuals tend to have particularly high-SES friends. These estimates of homophily are very similar (slope of 0.46 for full range, 1.02 between the 90th and 100th percentiles) when we restrict attention to an individual’s ten closest friends (defined based on the frequency of public interactions such as likes, tags, wall posts, and comments), showing that our estimates are not significantly affected by the strength of friendships or the number of Facebook friends that people have.

For our analyses below, it is useful to measure connections between individuals in different parts of the SES distribution. For simplicity, in our main analysis, we split individuals into two groups based on their SES: below-median and above-median SES (which we refer to below as “low SES” and “high SES,” respectively). On average, 38.8% of the friends of below-median-SES individuals have above-median SES, while 70.6% of the friends of above-median-SES individuals have above-median SES. Since 50% of individuals have above-median SES by definition, high-SES friends are *under-represented* by 22.4% ($1 - \frac{0.388}{0.5} = 0.224$) among low-SES individuals relative to their share in the population. Conversely, high-SES friends are *over-represented* by 41.2% among high-SES individuals ($\frac{0.706}{0.5} - 1 = 0.412$). Note that the share of high-SES friends for low- and high-SES individuals averages to 54.7% rather than 50% because high-SES people have more friends than low-SES people on average (Table 1).

If high-SES and low-SES individuals were to make friendships independent of socioeconomic status (i.e., there were no homophily by SES) and also were to make the same number of friends on average, then 50% of low-SES individuals’ friends would have high SES. In practice, above-median-SES individuals have 25.4% more friends than below-median-SES individuals on average (Table 1). If high-SES people continue to make 25.4% more friends than low-SES people, but friendships were formed independent of SES, the share of high-SES friends among low-SES individuals would be $\frac{1.254}{1+1.254} = 55.6\%$. Relative to that benchmark, low-SES individuals make 30.2% fewer high-SES friends than they would

in the absence of homophily.

We go beyond the two-group median split by examining connections between individuals in different deciles of the SES distribution. Table 1 presents a matrix of intra-decile friendship rates, showing the likelihood of friendship formation for people from different deciles of the SES distribution. Connectedness is lower between deciles that are further apart; for instance, top-decile friends are under-represented among people in the bottom decile by 75% relative to their population share ($1 - \frac{0.025}{0.1} = 0.75$), more than three times larger than the corresponding 22.4% under-representation of above-median friends among below-median individuals.

Childhood Economic Connectedness

In addition to measuring economic connectedness among adults, we use parent-child linkages to analyze economic connectedness based on the childhood friendships of individuals from different family backgrounds. Social capital during individuals' formative years may be particularly relevant for intergenerational income mobility (Loury 1977).

We measure childhood economic connectedness by analyzing homophily in high school friendships by parents' SES (see Section VI.D). Figure 1b plots the mean parental SES rank of a given individual's five closest high school friends against the SES rank of the individual's own parents. There is less homophily by parental SES during childhood than by own SES in adulthood, with a slope of 0.31 instead of 0.44. Much of this difference in slopes arises from the fact that the SES among friends from high school in adulthood is more similar than the SES of their parents, perhaps because children who befriend each other tend to follow similar trajectories (Chetty et al. 2022).

The series in squares in Figure 1b shows analogous estimates of homophily by parental SES rank among high school students using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a representative survey of high school students that contains self-reported information on close friendships (see Supplementary Information A.5.2). We obtain almost identical point estimates of homophily (slope = 0.31) by parental SES rank among high school friends in the Facebook and Add Health data. This comparison suggests that selection biases in Facebook usage or measurement error in the friendship links and SES ranks do not substantially distort our estimates of homophily.

Economic Connectedness Across Areas

The Facebook data, which is about 3,500 times larger than the Add Health sample, offers adequate precision and information to allow us to measure economic connectedness not just at the national level, but within specific communities such as a given neighborhood or school. We define the level of economic connectedness (EC) in a community as the average share of above-median-SES friends among below-median-SES members of that community divided by 50%, in order to quantify the average degree

of under-representation of high-SES friends among low-SES people (see Section VI.D for an algebraic definition). A value of 0 for 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. Although we focus on EC among *low-SES* individuals in particular, which we refer to simply as economic connectedness, we also construct and release analogous measures of community-level EC for high-SES individuals.

Figure 2a maps economic connectedness by county in the United States. Economic connectedness varies significantly across areas. Counties in the bottom decile (unweighted) of connectedness have ECs less than 0.58; that is, below-median-SES individuals have about 42% fewer above-median-SES friends than one would expect in the absence of homophily. Counties in the top decile have ECs of 1.05 or higher, approximately commensurate to what one would expect based on random sampling of friends from the national distribution, adjusting for the fact that high-SES people make more friends as discussed above. This geographic variation in connectedness is driven partly by differences in the fraction of high-SES individuals in an area and partly by differences in the rates at which low-SES individuals befriend high-SES individuals in their area; we decompose the relative contributions of these two factors, which we refer to as exposure and friending bias, in the next paper in this series (Chetty et al. 2022).

Economic connectedness is generally lowest in the Southeast, the Southwest, and industrial cities in the Midwest. It is highest in the rural Midwest and on the East Coast. The mean standard error of the county-level EC estimates is 0.004 (see Section VI.D and Supplementary Information B.3), implying that virtually all of the variation in Figure 2a reflects true differences in economic connectedness across areas rather than sampling error.

Economic connectedness varies not just across counties but also across neighborhoods within counties: 42% of the variation in EC across ZIP codes is within counties. Figure 2b illustrates this local variation by mapping EC by ZIP code (formally, ZIP code tabulation areas) in the Los Angeles metropolitan area (analogous maps for all ZIP codes in the U.S. are available at www.socialcapital.org). EC ranges from 0.62 to 1.25 between ZIP codes at the 10th and 90th percentiles of the EC distribution within the Los Angeles metro area (Los Angeles, Orange, and Ventura counties). Economic connectedness is lowest in the lowest-income neighborhoods of Los Angeles, such as Watts in central LA, where EC is 0.45. EC is generally higher in higher-income areas, but there is significant variation in EC even within those areas, with some places (such as Echo Park) having relatively low EC despite having many high-SES residents.

More broadly, looking outside Los Angeles, virtually none of the lowest-income ZIP codes in the U.S.

exhibit high levels of EC. It may be that there is little scope for low-SES people to connect with higher-SES people when there are few such people in the vicinity, echoing Blau’s observation that “persons cannot associate without having opportunities for contact” (Blau 1977). In our analysis, this point is an empirical result rather than a mechanical consequence of contact because low-SES individuals in low-income areas could in principle be friends with high-SES people outside their neighborhoods. In practice, such connections appear to be relatively rare. On the other hand, the presence of high-SES neighbors does not guarantee that low-SES people connect with those individuals: many higher-income neighborhoods still have EC well below 1.

The spatial patterns documented above are robust to the way in which economic connectedness is measured. For example, Supplementary Table 2 shows that we find similar spatial patterns for economic connectedness when restricting to individuals’ ten closest friends (correlation = 0.99 across counties). Similarly, the mean friend rank of individuals at the 25th percentile of the SES distribution, a measure that controls for differences in the SES distributions *within* the below-median and above-median groups, has an across-county correlation of 0.98 with our baseline EC measure. The share of top-quintile-SES friends among bottom-quintile-SES individuals in a county has a correlation of 0.74 with our baseline below- vs. above-median EC measure across counties. Childhood EC also exhibits broadly similar spatial patterns. Using two measures of childhood EC—one constructed for Facebook users from the SES of high school friends’ parents, and the other constructed for a sample of current 13-17 year-olds on Instagram (see Section VI.D)—we find across-county correlations of 0.61 and 0.82, respectively, with our baseline EC measure (see Supplementary Table 2). This last result suggests that differences in connectedness across areas are relatively stable over time, consistent with the high degree of serial correlation in our baseline county-level EC measure across birth cohorts (Supplementary Figure 1).

Connectedness by Other Attributes

We also measure connectedness between (i) individuals who do vs. do not use English as their primary language, and (ii) individuals between the ages of 25 and 34 vs. individuals between the ages of 35 and 44. Language and age connectedness exhibit different spatial patterns from economic connectedness (Figure 3). For example, the across-county correlation between language connectedness and economic connectedness is only 0.10 (Table 2). Hence, it is not simply that some areas exhibit high levels of connectedness across all types of individuals; rather, the degree of connectedness varies across different characteristics.

II.B Cohesiveness

Many theoretical studies have shown how the structure of social networks can shape a variety of outcomes, from human capital formation to the degree of adherence to social norms (Ballester et al. 2006; Coleman 1988; Jackson 2019). These studies of social capital conceptualize the *cohesiveness* of networks in two ways: (i) the cohesiveness of a given individual’s personal network (measured, for example, by the extent to which their friends are in turn friends with each other), and (ii) the cohesiveness of the whole community (measured by the degree of fragmentation into subcommunities). Empirical studies have shown that these measures are associated with a range of outcomes, including the dynamics of various types of contagion (Alatas et al. 2016; Centola 2018; Centola et al. 2007; Hill et al. 2010; Jackson and Storms 2018; Rand et al. 2011, 2014; Watts 2004). Motivated by this literature, we construct three measures of social capital that characterize the structure of friendship links in a community.

The first measure is clustering: the rate with which two friends of a given person are in turn friends with each other. The logic underlying clustering as a measure of social capital is that if a person’s friends are friends with each other, they can act together to pressure or sanction that person, which enforces norms and induces pro-social behavior and investment. Clustering ranges from 0 to 1, with a value of 0 meaning that all of a person’s friends are isolated from each other and 1 meaning that all of a person’s friends are friends with each other. We measure the degree of clustering in a community as the average rate of clustering in friendships for people living in that community (see Section VI.E).

A related, but distinct, measure of cohesiveness is the support ratio, which captures the rate at which pairs of friends in a community have other friends in common. The potential role of this measure of social capital can be microfounded in game theoretic models of the extent to which cooperative behavior between two individuals can be sustained: when two people have friends in common, their mutual friends can witness their behavior and react to it by enforcing norms (Jackson et al. 2012). We say that a friendship between two people is “supported” if they have at least one other friend in common. We measure the support ratio in a given community as the share of friendships among its members that are supported (see Section VI.E). A community’s support ratio varies from 0 to 1, with 0 implying that none of the friendships between members of a community are supported, and 1 implying that all such friendships are supported.

The third measure of network cohesiveness we consider is spectral homophily (Golub and Jackson 2012), which captures the extent to which a network is fragmented into separate groups (see Section VI.E for a formal definition). Spectral homophily also ranges from 0 to 1: a value of 0 implies no homophily, such that individuals are equally likely to be friends with any other member of the community, while a value of 1 implies that the network fragments into two or more distinct groups across which no one

interacts.

All three of these measures of network cohesiveness exhibit broadly similar (though not identical) spatial patterns, with absolute pairwise correlations of 0.51–0.64 with each other across counties (Table 2). In general, clustering and support ratios are highest in the South, Appalachia, and rural Midwest (see Figure 2c and Figure 3c). Spectral homophily tends to be lowest in these areas and highest in the Southwest (see Figure 3d). Dense urban centers often exhibit high levels of spectral homophily and low levels of clustering, consistent with Coleman’s prediction that areas with greater levels of geographic mobility will have less clustered networks (Coleman 1988).

The network cohesiveness measures exhibit different geographic patterns from economic connectedness, with correlations ranging from -0.25 to 0.01 with economic connectedness across counties (Table 2 and Panels A and C of Figure 2). These differences emerge not just across counties but across neighborhoods within counties, as illustrated by the ZIP code-level maps of the Los Angeles metro area in Panels B and D of Figure 2.

II.C Civic Engagement

A third widely discussed concept of social capital is based on levels of civic engagement and pro-social behavior rather than on the structure of networks (Putnam 1995; Simmel 1902; Thomas and Znaniecki 1919). This form of social capital has been measured using self-reported levels of trust, rates of volunteering, or rates of membership in local organizations (Glaeser et al. 2000; Rupasingha et al. 2006; Social Capital Project 2018). Such measures are often associated with various outcomes across regions and countries, ranging from economic growth to political accountability (Banfield 1958; Knack and Keefer 1997; Nannicini et al. 2013; Putnam 2000; Tabellini 2010).

Because they do not rely on network data, state- and county-level indices of civic engagement based on survey data are widely available. Here, we build upon prior efforts by constructing measures of civic engagement at the more granular ZIP code level, taking advantage of the large sample sizes available in the Facebook data.

One common way to measure civic engagement is based on rates of volunteering (Social Capital Project 2018). Building on prior work by Herdağdelen et al. (2021), we construct a proxy for the rate of volunteering in an area based on the fraction of Facebook users in that area who are members of at least one “volunteering” or “activism” group as classified based on their titles. Such groups include, for example, Neighbors Helping Neighbors or Adopt a Senior; see Section VI.F for details. This measure has a population-weighted correlation of 0.58 with survey-based measures of volunteering rates across states from the Social Capital Project (2018), suggesting that it captures a similar concept.

Another prominent measure of civic engagement is the density of civic organizations in a county (Rupasingha et al. 2006). We construct a granular measure of the density of civic organizations (e.g., non-profits) based on the number of Facebook pages for such organizations in an area divided by its population (see Section VI.E for details). Our index has a population-weighted correlation of 0.67 with the Penn State Index (Rupasingha et al. 2006) across counties (Table 2).

Our two measures of civic engagement vary substantially across areas and exhibit somewhat similar geographic patterns, with a population-weighted correlation of 0.46 across counties. Rates of volunteering are highest in the Pacific Northwest and lowest in the Southeast (Figure 2e). Civic organizations are most common in the Rocky Mountains, Pacific Northwest, and New England, and least common in parts of the South (Figure 3e). Both measures of civic engagement also vary substantially across ZIP codes within counties (Figure 2f).

Civic engagement is positively correlated both with measures of network cohesiveness and measures of economic connectedness (see Table 2). Most notably, volunteering rates have a correlation of 0.46 with EC across counties.

In summary, the new measures of social capital constructed here underscore the importance of specifying a particular notion of social capital when assessing a community’s level of social capital. This result is in line with prior observations based on ethnographic and theoretical analyses (DeFilippis 2001; Jackson 2020; Portes 1998) that have illustrated how a single community can exhibit different levels of social capital depending on the concept being measured. For example, Portes (1998) notes that “since the publication of Stack (1974), sociologists know that everyday survival in poor urban communities frequently depends on close interaction with kin and friends in similar situations. The problem is that such ties seldom reach beyond the inner city, thus depriving their inhabitants of sources of information about employment opportunities elsewhere and ways to attain them.” Our quantitative measures confirm these ethnographic observations in specific communities on a national scale, showing, for example, that poor urban communities with very cohesive networks often do not provide connections to high-SES individuals.

The benefit of having measures of social capital for all communities in the U.S. is that they can be used to study which types of social capital matter for various outcomes of interest. In the next section, we investigate which forms of social capital are associated with one prominent outcome that many have hypothesized to rely on social capital: upward economic mobility.

III Social Capital and Upward Income Mobility

Rates of upward income mobility—children’s chances of rising up in the income distribution conditional on growing up in low-income families—vary substantially across areas in the United States (Chetty et al. 2014). A large literature has sought to understand and explain these differences. One widely-discussed hypothesis, based on indirect proxies and ethnographic evidence, is that differences in economic mobility across areas may be related to differences in social capital (Brooks 2020; Putnam 2016; Social Capital Project 2019).

In this section, we study this hypothesis by analyzing the associations between the measures of social capital constructed above and economic mobility across areas. We obtain statistics on intergenerational income mobility and other related outcomes, such as high school graduation rates and teenage birth rates, from the publicly available Opportunity Atlas, which constructs these statistics based on Census and tax data covering all children born in the U.S. between 1978–83 (Chetty et al. 2018). We focus on correlations between upward mobility and social capital across areas rather than individuals because area-level variation is arguably more likely to be driven by institutional, policy-relevant factors than individual-level variation. Furthermore, we have very precise measures of economic mobility (constructed using tax data) at the area level; at the individual level, estimates of income mobility using Facebook data have greater measurement error, potentially inflating correlations between one’s own outcomes and friends’ SES.

We begin by examining correlations between social capital and economic mobility across counties and then turn to a more granular ZIP code-level analysis.

County-Level Correlations

Figure 4a reports univariate correlations (weighted by the number of children with below-national-median parental income) across counties between each measure of social capital constructed above and upward income mobility (see also Table 3). We define upward income mobility in each county as the average income percentile rank in adulthood of children who grew up in that county with parents at the 25th percentile of the national parental household income distribution (Chetty et al. 2018). Economic connectedness is strongly positively correlated with upward income mobility (correlation = 0.65, s.e. = 0.04), whereas all the other measures of social capital are not strongly related to mobility.

Figure 5 shows the relationship between economic connectedness and mobility non-parametrically by presenting a scatter plot of upward income mobility vs. EC for the 200 most populous counties. Children who grow up in counties where low-SES individuals have more high-SES friends tend to have much higher rates of upward mobility. As an example, low-SES individuals have a much larger fraction of high-SES friends in Minneapolis (49%, corresponding to EC of 0.98) compared to Indianapolis (32%,

EC of 0.65). Correspondingly, children who grow up in low-income families have much higher incomes in adulthood in Minneapolis (on average reaching the 43rd percentile of the household income distribution for children at age 35, roughly \$34,300 in 2015) than in Indianapolis (the 34th percentile, or \$24,700).

On average, an increase in EC of 0.5 units (equivalent to raising the share of high-SES friends among low-SES people from 25% to 50%, and approximately equal to the difference in EC between the 10th and 90th percentile counties) is associated with an 8.2 percentile increase in children's incomes in adulthood. This is a large difference: for context, note that children with high-income (above-median) parents end up 17 percentiles higher in the household income distribution on average than children with low-income (below-median) parents (Figure 6). We find similarly strong associations between economic connectedness and many other outcomes related to social mobility, such as high school completion rates and teenage birth rates (Figure 7).

Returning to Figure 4a, other measures of connectedness across groups—between non-English and English speakers or between younger and older individuals—are less strongly associated with upward mobility. Communities with greater connectedness across groups in general do not necessarily have higher levels of upward mobility; connections across class lines are what appear to matter.

Measures of network cohesion (e.g., clustering and support ratios) also do not correlate strongly with observational measures of upward mobility. This is because there are many areas that exhibit highly cohesive networks—and thus might be thought of as tightly-knit communities—but that nevertheless have low levels of economic connectedness and correspondingly low levels of upward mobility. One potential explanation for this pattern is that while those communities have strong social connections among their predominantly low-income residents (bonding social capital), they are not well connected to individuals from higher-SES backgrounds who can provide the types of resources, opportunities, and information (Bourdieu 1986; Loury 1977) needed to rise economically (bridging social capital).

Finally, we examine associations between economic mobility and measures of civic engagement. The widely used Penn State Index of participation in civic organizations has a correlation of 0.06 across counties with upward mobility (Rupasingha et al. 2006). We find similarly weak associations of upward mobility with the density of civic organizations and volunteering rates. The difference between these findings and prior work that finds stronger associations between civic engagement and economic mobility is primarily because we weight our correlations by the number of children with below-national-median parental income. As a result, rural areas—where civic engagement is more strongly correlated with mobility—receive less weight in our correlations (see Supplementary Information C.2).

When we regress standardized measures of upward mobility on standardized versions of all of the social capital measures together, economic connectedness remains the strongest predictor of upward

mobility by a significant margin, and measures of civic engagement and network cohesiveness have coefficients near zero (Figure 4b). Furthermore, a LASSO regression chooses EC as the first social capital measure to include and places greater weight on EC than on other measures (Supplementary Figure 2a), and the incremental R-squared of including EC conditional on all the other social capital measures is an order of magnitude larger than the incremental R-squared of including any of the other measures (Supplementary Figure 2c).

ZIP Code-Level Correlations

When studying variation across ZIP codes instead of counties in the U.S., we find very similar correlations between upward income mobility and social capital measures (Table 3, Supplementary Figure 3). In particular, upward mobility is highly correlated with economic connectedness across ZIP codes (correlation = 0.69, Supplementary Figure 4), but more weakly correlated with all the other social capital measures. Going from the 10th to the 90th percentile ZIP code in the U.S. in terms of economic connectedness is associated with an 11 percentile increase in the mean rank of children growing up in low-income families—comparable to the difference in upward mobility between Black and white children with low-income parents of 12.6 percentiles (Chetty et al. 2020).

Next, we examine the association between social capital measures and mobility across ZIP codes *within* the same county to assess whether the ZIP code-level relationships differ across counties. The ZIP code-level correlation between economic connectedness and mobility is strongly positive within virtually all counties. On the other hand, there is substantial heterogeneity in the ZIP code-level relationships between other measures of social capital and mobility across counties. Figure 8a illustrates this by presenting binned scatter plots of the relationship between upward mobility and clustering coefficients by ZIP code across four cities in Ohio: Akron, Cleveland, Columbus, and Youngstown. In Cleveland and Columbus, where baseline levels of clustering are relatively low, neighborhoods with higher clustering coefficients have significantly higher levels of upward income mobility. But in Akron and Youngstown, which generally have higher levels of clustering, clustering and upward mobility are *negatively* correlated. Hence, it is not that clustering coefficients have no signal in predicting economic mobility; rather, their relationship with mobility varies across places, in part depending on their average levels of clustering.

The relationship between economic connectedness and mobility is much more stable across the same four cities, as shown in Figure 8b. The relationships between clustering coefficients and economic connectedness closely match those for mobility: in Cleveland and Columbus, clustering coefficients and economic connectedness are positively related, while in Akron and Youngstown, they are negatively related (Figure 8c). More generally, building upon these examples, we find that clustering is often positively correlated with economic connectedness and mobility when clustering is low, whereas it is

generally negatively correlated with both EC and mobility when levels of clustering are high. These patterns suggest that economic connectedness may mediate the relationship between the other social capital measures and mobility; that is, any observed links between other social capital measures and mobility might run through economic connectedness.

Figure 8d generalizes the four examples by plotting the distributions of the correlations between upward mobility and various measures of social capital across ZIP codes within the 250 most populous U.S. counties. For economic connectedness, the distribution is sharply peaked around 0.7: economic connectedness and mobility are positively correlated across ZIP codes in virtually all counties. In contrast, the other social capital measures exhibit more diffuse distributions across counties. Importantly, these differences are not just due to sampling error in the correlations; adjusting for noise by calculating the reliability of the estimates and the standard deviation of the latent signal distribution yields similar conclusions (see Supplementary Table 3).

To summarize, measures of social capital that are based solely on the structure of the network graph (network cohesion) or purely on individuals’ civic behaviors (civic engagement) do not have robust associations with observational measures of economic mobility across areas. Measures that combine data on networks with information on socioeconomic status have stronger and more stable relationships with economic mobility.

Having established that economic connectedness stands out among social capital measures as a strong predictor of economic mobility, in the rest of the paper we focus on understanding the source of this correlation: why do more economically connected areas tend to have higher rates of economic mobility?

IV Why is Connectedness Related to Mobility?

There are many theories for why economic connectedness could have a positive causal effect on mobility. For example, economic mobility might be facilitated by connections to people who can shape aspirations or provide access to information and job opportunities (Lin and Dumin 1986; Loury 1977). This interpretation is consistent with the argument (Putnam 2000) that bridging capital—a concept that encompasses economic connectedness—is particularly valuable for “getting ahead.” However, there are also many alternative explanations for the correlation between economic connectedness and mobility that do not rely on a causal effect of connectedness on mobility. We evaluate three such possibilities in turn—reverse causality, selection effects, and omitted variables—with the broader aim of better understanding the channels through which connectedness and mobility are related.

Reverse Causality

The first alternative explanation for the correlation between connectedness and mobility we consider is *reverse causality*: greater economic mobility could lead to greater economic connectedness. Specifically, in our baseline analysis, we correlate rates of upward income mobility with economic connectedness measured among adults. Because friendships and SES are measured in adulthood, economic connectedness may itself be influenced by rates of intergenerational mobility. For example, in high-upward-mobility places, many children from low-SES families have high incomes as adults and may retain friendships with individuals who remain low-SES. This would lead high-mobility areas to have a high rate of friendships between people with different SES in adulthood, even absent any effect of connectedness on mobility.

To assess the importance of reverse causality, we examine the association between economic mobility and *childhood* EC, based on childhood friendships and parental SES. Because childhood friendships are made before people begin working, they cannot be directly influenced by rates of economic mobility. We measure childhood EC using two sources of data, each of which has certain benefits and drawbacks (see Section VI.E). The first is based on the high school friends and parental SES of individuals in our primary Facebook analysis sample. The second uses data from Instagram for individuals aged 13-17 in 2022, measuring parental SES based on the teenagers' residential ZIP codes and phone models.

The correlation between upward mobility and childhood EC across counties remains high with both of these measures: 0.44 using parental SES in the Facebook data and 0.62 using the Instagram data (Table 3). The fact that upward mobility remains strongly correlated with childhood EC implies that causal effects of mobility on connectedness account for at most a small portion of the correlation between the two variables.

Causal Effects of Place vs. Selection

A second potential non-causal explanation for the link between economic connectedness and mobility is *selection*: the types of families who live in high-EC areas may inherently have higher rates of mobility (e.g., because they have more education or wealth), independent of where they live. For example, the types of low-income families who choose to live in high-EC areas may have demographic characteristics or make other choices that increase their children's rates of upward mobility even absent any causal effect of EC on outcomes.

One of the most salient forms of residential sorting in the U.S. is segregation by race and ethnicity. Such segregation could induce a correlation between EC and mobility. For example, areas with larger Black populations tend to have lower levels of economic connectedness (Supplementary Table 4). Since Black Americans have lower rates of upward mobility than white Americans (Chetty et al. 2020) — which could be due to factors unrelated to differences in EC such as discrimination— differences in

racial composition across neighborhoods could induce a spurious association between EC and mobility when pooling across races.

The simplest way of assessing the importance of differences by race would be to replicate our baseline correlations conditioning on race, for instance by correlating upward mobility and connectedness among Black individuals. As a feasible alternative in the absence of individual-level data on race, we focus on racially homogeneous areas: counties or ZIP codes where most of the residents are of the same race (based on publicly available data from the Census). We then correlate race-specific measures of economic mobility (Chetty et al. 2018) with EC (pooling all racial groups) within these racially homogeneous areas.

Table 4 reports the results of this analysis. Column 1 shows that the correlation between upward mobility for white individuals and overall EC is 0.68 in counties where at least 80% of residents are white (which have a mean white share of 90%). The correlation is similar (0.69) in counties where at least 90% of residents are white, and the mean white Share is 95% (Column 2). Columns 3 and 4 show that results are similar at the ZIP code level: in ZIP codes where at least 90% of residents are white (where 95% of people are white overall), the correlation between upward mobility and EC is 0.69. Columns 5 and 6 show similarly strong correlations between upward mobility for Black people and EC in predominantly Black ZIP codes. Columns 7 and 8 show somewhat smaller (though not statistically distinguishable) correlations between upward mobility for Hispanic people and EC in the few predominantly Hispanic ZIP codes. Note that we can only perform this analysis at the ZIP code level for Hispanic and Black individuals because there are very few counties that have more than 80% Black or Hispanic residents.

The results in Table 4 show that economic connectedness remains highly correlated with economic mobility even conditional on race, implying that sorting by race is unlikely to be the primary driver of the observed correlation between EC and mobility overall. Relationships between mobility and other measures of social capital also remain similar when restricting the sample to racially homogeneous areas (Supplementary Figure 5).

Of course, there are many dimensions beyond race on which families may sort across neighborhoods, such as their underlying human capital or their propensity to invest in their children’s education. To test for sorting on such dimensions, many of which are unobservable, one would ideally randomly assign families to low- vs. high-EC areas—thereby ensuring that families in high- and low-EC areas are comparable—and examine whether their children’s outcomes differ in adulthood. We approximate this experiment using quasi-experimental estimates of the causal effect of growing up for an additional year in each county in the U.S. on household income in adulthood from Chetty and Hendren (2018b), who

exploit variation in the age at which children move across counties to identify the causal effect of growing up in each county for children with parents at the 25th percentile of the income distribution. Under the identification assumption that the *timing* of moves is unrelated to children’s potential outcomes—an assumption validated in a series of experimental and quasi-experimental studies (Chetty and Hendren 2018a; Chetty et al. 2016a; Chyn 2018; Deutscher 2020)—differences in income in adulthood for children who move at younger vs. older ages to a given county reveal its causal effect on economic mobility.

We analyze the relationship between counties’ *causal* effects on upward mobility and economic connectedness in Figure 9. Following Chetty and Hendren (2018b), we measure each county’s causal effect as the mean change in an individual’s percentile income rank from growing up from birth (for 20 years) in that county instead of the average county in the U.S. The figure presents a binned scatter plot of counties’ causal effects on upward mobility vs. their EC. Higher-EC counties have larger causal effects on upward mobility, with a correlation of 0.44 (s.e. = 0.06) after correcting for sampling error in the causal effect estimates (see Section VI.G). In a multivariable regression of counties’ causal mobility effects on all our social capital measures, economic connectedness remains highly correlated with causal effects on mobility, while most other social capital measures do not exhibit statistically significant associations (Supplementary Figure 6).

The slope of the relationship in Figure 9 implies that growing up from birth in a county with 1 unit higher EC increases income in adulthood by 9.8 percentiles (a 30.7% increase relative to mean income ranks) for children with low-income parents. This estimate implies that moving at birth from the 10th to 90th percentile ZIP code in terms of EC—a move associated with an increase in EC of 0.57—would increase children’s household income in adulthood by 17.5% on average. As another benchmark, note that the average difference in EC between low- and high-SES individuals is 0.636. If low-SES children were to grow up in counties with EC comparable to the average high-SES child, their incomes would increase on average by $0.636 \times 30.7 = 19.5\%$ (equivalent to 6.23 percentiles). This increase in income would close about 37% of the current 17 percentile gap in income in adulthood between children with parents at the 25th vs. the 75th percentiles of the income distribution.

We conclude that the correlation between EC and mobility is not driven purely by differences in the types of families who live in high EC areas; instead, growing up in an area with higher EC *causes* significantly higher rates of upward mobility.

Connectedness vs. Other Factors

Higher-EC areas may generate higher levels of mobility for two reasons: either economic connectedness itself has a causal effect on mobility or high-EC places have other characteristics (e.g., better schools) that generate higher levels of mobility. As a step toward distinguishing these two explanations,

we compare the relative explanatory power of EC and the strongest neighborhood-level predictors of economic mobility identified in prior work.

We begin by analyzing incomes across neighborhoods. Several studies have shown that areas with lower incomes and more highly concentrated poverty have lower rates of economic mobility (Chetty et al. 2016b; Manduca and Sampson 2019). Motivated by such findings, many place-based policies use high poverty rates as a marker to identify “low opportunity” neighborhoods that are eligible for special tax credits and resources, and recent work has sought to help families move to lower-poverty neighborhoods to improve their economic prospects (Bergman et al. 2019).

Figure 10a shows univariate county-level correlations between upward mobility and measures of income as well as various other neighborhood characteristics (results at the ZIP code level, shown in Supplementary Figure 4b, are similar). The share of individuals above the poverty line and median household incomes have correlations of 0.3-0.35 with economic mobility across counties. When we regress upward income mobility on both economic connectedness and measures of local income levels (poverty rates or median household incomes), connectedness remains a strong predictor of upward mobility whereas measures of local income levels lose much of their predictive power at both the county and ZIP code levels (Table 5a; Supplementary Figures 7 and 8).

These findings suggest that economic connectedness may be a mediator through which concentrated poverty affects upward mobility: living in a lower-income neighborhood may inhibit upward mobility insofar as it reduces interaction with higher-SES people, but does not appear to have a strong influence beyond its effect on economic connectedness. Figure 11 demonstrates this point more directly by presenting a scatter plot of economic connectedness vs. median household income by ZIP code. The dots are colored based on the level of upward income mobility for children who grew up in low-income families in that ZIP code, with blue representing areas with higher levels of upward mobility and red representing areas with lower levels of mobility. Horizontal slices of the graph—neighborhoods with different levels of median income but comparable levels of economic connectedness—tend to have similar levels of economic mobility. By contrast, vertical slices of the graph—areas with comparable incomes, but different levels of EC—transition from low to high economic mobility as EC rises. These results imply that it is growing up in an area with high EC—rather than just around high-income people—that leads to better prospects for upward mobility.

While local income levels explain little of the relationship between economic connectedness and outcomes for children starting out in low-income (25th percentile) families, they do appear to mediate the relationship between connectedness and outcomes for children in *high-income* (75th percentile) families. We illustrate this result in Figure 12a. As a reference, the series in orange circles presents a

binned scatter plot of upward mobility vs. economic connectedness for low-SES individuals, by county. This series is similar to the scatter plot in Figure 5, except that we include all counties and group them into twenty equal-sized bins based on their level of economic connectedness to show the conditional expectation of upward mobility given economic connectedness non-parametrically. Consistent with the pattern in Figure 5, we find a strong positive slope of 18.2.

Now consider the relationship between the average income ranks in adulthood of children with parents at the 75th percentile and the fraction of low-SES friends that high-SES individuals have. This relationship (plotted in blue circles) is flatter than that for low-SES individuals. A 1 unit increase in cross-group connectedness—defined here as twice the fraction of low-SES friends among high-SES individuals—is associated with an 8.6 percentile reduction in mean income rank for children with parents at the 75th percentile. Importantly, after controlling for the fraction of high-SES individuals in the county, greater cross-group economic connectedness remains strongly positively associated with outcomes for children with parents at the 25th percentile (as established above), but is now uncorrelated with the economic outcomes for children with parents at the 75th percentile (Figure 12b). One potential explanation for this pattern is that greater *interaction* between low- and high-SES households conditional on the income mix in an area benefits low-SES individuals without harming high-SES individuals; however, greater income mixing (integration) benefits low-SES individuals partly at the expense of high-SES individuals by redistributing public goods (e.g., local public school funding) from people with higher incomes to people with lower incomes. These results raise the possibility that more economically connected communities can benefit the poor with limited adverse impacts on the rich, particularly if increasing cross-SES connections does not require changing the income mix or resources in an area.

Going beyond average income levels, prior research has also shown that more segregated counties—where people of different incomes or racial backgrounds live in separate neighborhoods—tend to exhibit lower levels of economic mobility. Indices of segregation by income and race (constructed from Census data using the method of Reardon and Bischoff (2011), see Supplementary Information A.5.1) have negative correlations of 0.17–0.21 with economic mobility across counties, significantly lower than the correlation of 0.65 observed with economic connectedness. Hence, using network data to directly measure interaction (rather than proxying for it based on residential location) adds considerable explanatory power for understanding economic mobility. Moreover, when we regress upward mobility on both economic connectedness and segregation measures, connectedness remains a strong predictor of upward mobility whereas segregation indices lose their predictive power (Table 5b, Supplementary Figure 9).

Prior work has established that Black individuals living in neighborhoods with larger Black popula-

tion shares have poorer educational and economic outcomes on average (Cutler and Glaeser 1997). We replicate these results in the odd-numbered columns of Table 5c by regressing upward income mobility for Black and white individuals on the share of Black residents in an area (for both counties and ZIP codes). The corresponding even-numbered columns show that controlling for economic connectedness eliminates or even reverses the relationship between the share of Black residents and rates of upward mobility (see also Supplementary Figure 10). Intuitively, areas with larger Black shares tend to have lower levels of EC (Supplementary Table 4), and this relationship accounts for the negative correlation between Black shares and rates of mobility.

Prior research has also found a strong negative correlation between income inequality within a generation (measured for example by the Gini coefficient) and upward mobility across generations, coined the “Great Gatsby curve” (Corak 2013; Krueger 2012). Controlling for economic connectedness essentially eliminates this relationship (Columns 5 and 6 of Table 5b, Supplementary Figure 9). Greater income inequality is associated with less economic connectedness, and that relationship can largely explain the negative correlation between inequality and mobility. In short, a lack of economic connectedness may be a key reason that upward mobility is lower in areas with larger Black populations and greater inequality (Durlauf and Seshadri 2017).

Finally, we turn to other factors that have been explored in prior work, ranging from the quality of local schools to job availability to measures of family structure. Economic connectedness is more strongly correlated with upward economic mobility than almost all of those characteristics in univariate specifications (Figure 10a). In a multivariable regression of upward mobility on EC along with other predictors that have the highest univariate correlations with mobility, economic connectedness is the strongest predictor of upward mobility (Figure 10b) and has the largest incremental R-squared (Supplementary Figure 2d). Economic connectedness is also among the first variables—along with single parent shares—that are chosen by a LASSO regression as predictors of economic mobility (Supplementary Figure 2b).

In summary, places with higher levels of economic connectedness generate higher levels of economic mobility even controlling for the strongest neighborhood-level predictors of economic mobility identified to date. Moreover, the relationships between these other neighborhood characteristics and mobility become much weaker once we control for EC, suggesting that the links between those factors and mobility may run through their impacts on economic connectedness. These findings suggest that other observable neighborhood characteristics do not explain why higher-EC areas generate higher levels of upward mobility, calling for further focus on causal mechanisms through which economic connectedness itself may affect mobility.

V Discussion

Measuring social capital has proven to be more challenging than measuring other forms of capital, such as financial or human capital. Data from online social networking platforms offer a path to solving this problem. The new measures of social capital constructed here provide a rich picture of how social capital varies across areas in the United States. Different notions of social capital—connectedness across socioeconomic lines, the cohesiveness of a community, and civic engagement—exhibit very different spatial patterns. Many communities are rich in one form of social capital but poor in others.

Distinguishing these forms of social capital is important because some types of social capital are more correlated with certain outcomes than others. For instance, economic connectedness—the share of high-SES friends among low-SES people—is strongly associated with upward income mobility, whereas other forms of social capital are not. Areas with higher economic connectedness have large positive causal effects on children’s prospects for upward mobility. We caution, however, that this finding does not imply that economic connectedness is the best or most important measure of social capital in general. Economic connectedness may be the best predictor of economic mobility because mobility is essentially a measure of the degree to which individuals can increase their own SES, making it natural that links to higher-SES individuals are related to that outcome—consistent with hypothesis that bridging capital is useful specifically for “getting ahead” (rather than simply “getting by”) (Lancee 2010; Putnam 2000). For other outcomes, other social capital indices that we have constructed here may be stronger predictors. For example, differences in life expectancy among low-income individuals across counties are more strongly predicted by network cohesiveness measures (clustering coefficients and support ratios) than economic connectedness (Supplementary Figure 11, Supplementary Information C.3).

Our analysis raises three sets of questions for future research. First, it would be useful to conduct systematic studies of the forms of social capital that matter for other outcomes. Which forms of social capital matter for health behaviors or the formation of political preferences? Do interactions between different types of social capital matter for certain outcomes? The publicly available statistics constructed here can be used to study many such questions.

Second, it would be valuable to build on the methods developed here and construct analogous measures of social capital beyond the United States, either using social network data or other sources of network information such as financial transactions or mobile phone data (Kuchler and Stroebel 2022). While many of the lessons obtained from our analysis of the U.S. are likely to generalize more broadly, international comparisons would enrich our understanding of social capital and its determinants.

Finally, it would be useful to directly study whether efforts to increase economic connectedness can increase intergenerational income mobility. Doing so requires an understanding of the determinants of

economic connectedness and potential interventions to increase it. We turn to these questions in the next paper in this series, where we study why economic connectedness varies with socioeconomic status and how we can increase connectedness among low-SES individuals (Chetty et al. 2022).

VI Data and Methods

VI.A Sample Construction

This section describes the methods used to generate the data analysed 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.H).

We work with privacy-protected data from Facebook. Survey data show that more than 69% of the U.S. adult population used Facebook in 2019, and about three-quarters of those individuals did so every day (Perrin and Anderson 2019). The same survey also found that Facebook usage rates are fairly similar across income groups, education levels, and racial groups, as well as among urban, rural, and suburban residents; they are lower among older adults and slightly higher among women than men.

Starting from the raw Facebook data as of May 28, 2022, our primary analysis sample was constructed by limiting the data to users aged between 25 and 44 who resided in the United States, were active on the Facebook platform at least once in the prior 30 days, had at least 100 U.S.-based Facebook friends, and had a non-missing ZIP code. Our final analysis sample consists of 72.2 million Facebook users, who comprise 84% of the U.S. population between ages 25–44 (based on a comparison to the 2014–18 American Community Survey). We focus on the 25–44 age range because its Facebook usage rate is above 80%, higher than for other age groups, according to Perrin and Anderson (2019). In addition, the American Community Survey (ACS) publicly releases demographic data for certain age groups, one of which is ages 25–44, allowing us to compare our sample with the full population as well as to use ACS aggregates to predict socioeconomic status (see Section VI.B).

We do not link any external individual-level information to the Facebook data. However, we use various publicly available sources of aggregate statistics to supplement our analysis, including data on median incomes by block group from the 2014–18 ACS; data on economic mobility by Census tract and county from the Opportunity Atlas (Chetty et al. 2018); and measures of county- and ZIP-level characteristics such as racial shares and single parent shares from the ACS, the Census, and Chetty et al. (2014). We describe those data in detail in Supplementary Information A.5.

VI.B Variable Definitions

We construct the following sets of variables for each person in our analysis sample. We measured these variables on May 28, 2022.

Friendship Links. The data contain information on all friendship links between Facebook users. We focus only on friendships within our analysis sample; i.e., we exclude friendships with people below age 25 or above age 44, people who live outside the United States, or people who do not satisfy one of our other criteria for inclusion in the analysis sample.

Facebook friendship links need to be confirmed by both parties, and most Facebook friendship links are between individuals who have interacted in person (Jones et al. 2013). The Facebook friendship network can therefore be interpreted as providing data on people’s “real-world” friends and acquaintances rather than purely online connections. Because individuals tend to have many more friends on Facebook than they interact with regularly, we also verify that our results hold when focusing on an individual’s ten closest friends, where closeness is measured based on the frequency of public interactions such as likes, tags, wall posts, and comments.

Locations. Following Maas et al. (2019), we use location data to construct statistics at various geographic levels. Every individual is assigned a residential ZIP code and county based on information and activity on Facebook, including the city reported on Facebook profiles 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. 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. We construct our baseline SES measure in three steps, which are described in greater detail in Supplementary Information B.1.

First, for Facebook users who have Location History (LH) settings enabled, we use the ACS to collect the median household income in their Census block group. 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 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 listed in Supplementary Table 5. 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 turn out to be highly correlated with external measures of income across subgroups.

The algorithm described above is one of many potential ways of combining a set of underlying proxies for SES into a single measure. To verify that our findings are not sensitive to the specific variables or algorithm used to predict SES, we show that results are similar if we use a simple unweighted average of z-scores of the underlying proxies or if we directly use ZIP code median household incomes for all users, eschewing the prediction model and other proxies entirely (Supplementary Table 6).

Parental Socioeconomic Status. We link individuals in our primary analysis sample 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 Supplementary Information A.2 for details). We then use the SES of parents, constructed via the algorithm described above, to assign parental SES to individuals. Finally, we assign individuals a parental SES rank based on their predicted parental SES, ranking individuals based on parental SES relative to others in the same birth cohort. We are able to assign parental SES ranks for 31% of the individuals in our primary analysis sample.

High School Friendships. To identify friendships made in high school, we first use self-reports to assign individuals to schools. For people who do not report a high school, we use data on their friendship networks to impute those groups (see Supplementary Information A.3 for details). For the 3.3% of users who report multiple high schools, we select the school in which the user has the largest number of high school friends. This process yields information on high schools for 74.9% of individuals in our analysis sample. Finally, if an individual and one of their friends attended the same high school within three cohorts of each other, we identify them as high school friends.

VI.C Summary Statistics and Benchmarking

Table 6a shows summary statistics for our baseline sample and, for comparison, for those between ages 25-44 in the 2014-18 ACS. The Facebook sample is similar to the full population in terms of age, sex, and language. Consistent with prior work using Facebook data, women are slightly over-represented in our Facebook sample (53.6%) relative to men (Bailey et al. 2020b). The median individual in our analysis sample has 382 in-sample Facebook friends; in total, there are just under 21 billion friendship pairs between individuals in the sample.

Since much of our analysis relies on variation across areas, it is important that our sample has good coverage not just nationally but also across locations. In Supplementary Information A.1, we show that our sample has high coverage rates across the United States, and that coverage rates do not vary systematically across locations with different income levels or demographic characteristics.

Most of our analysis draws upon the SES measure constructed as described in the previous subsection. We evaluate the accuracy of this SES measure by correlating the fraction of households with above-median income within each ZIP code from the ACS with the estimated proportion of Facebook users with above-median SES in our sample. We find a (population-weighted) correlation of 0.88 between our estimates of the fraction of high-SES individuals and the ACS estimates at the ZIP code level. Furthermore, we find similarly high correlations between our estimates of the share of high-SES households and corresponding statistics drawn from external publicly available administrative datasets at the high school and college levels (see Chetty et al. 2022 for details).

For some parts of our analysis—in particular, for computing measures of economic connectedness during childhood—we focus on the subsample of individuals whom we can link to parents with an SES prediction and whom we can assign a high school based on self-reports and network-based imputations. Panel B of Table 6 presents summary statistics for this subsample of 19.4 million users, or about 27% of the full analysis sample. The characteristics of this subsample are broadly similar to those of the full sample, although users whom we can link to high schools and parents with SES predictions are about two years younger on average than users in the full sample, in large part because our approach does not allow us to assign SES predictions for parents older than 65. County-level median household incomes differ by \$876 between the samples, about 6% of a standard deviation.

We further evaluate our SES measure and parental linkages by comparing estimates of intergenerational economic mobility using our SES proxies to publicly available estimates based directly on household incomes from population-level tax data. We find a linear relationship between individuals' SES ranks and their parents' SES ranks across the distribution of parental SES, with a slope of 0.32 (Figure 6). This relationship is very similar to the estimated slope of 0.34 in population tax data

(Chetty et al. 2014), supporting the validity of both our SES imputations and parental linkages.

We conclude that our Facebook analysis samples are representative of the populations we seek to study and that our measures of socioeconomic status are aligned with external data.

VI.D Measuring Connectedness

Economic Connectedness. Let

$$f_{Q,i} \equiv \frac{[\text{Number of friends in SES quantile } Q]_i}{\text{Total number of friends}_i} \quad (1)$$

denote individual i 's share of friends from SES quantile Q . To obtain measures of the degree of homophily that are not sensitive to the size of each quantile bin, we normalize $f_{Q,i}$ by the share of individuals in the sample who belong to quantile Q , w_Q (e.g., $w_Q = 0.1$ for deciles). We then define person i 's *individual economic connectedness (IEC)* to individuals from quantile Q as

$$IEC_{Q,i} \equiv \frac{f_{Q,i}}{w_Q}. \quad (2)$$

We define the level of economic connectedness (EC) in community (county or ZIP code) c as the mean level of individual economic connectedness of low-SES (e.g., below-median) members of that community:

$$EC_c = \frac{\sum_{i \in L \cap c} IEC_i}{N_{Lc}}, \quad (3)$$

where N_{Lc} is the number of low-SES individuals in community c . When defining EC in a given community, we continue to rank individuals in the *national* SES distribution includes friendships to individuals residing outside that community. In the presence of homophily, EC ranges from 0 to 1, with a value of 1 indicating, for example, that half of below-median-SES individuals' friends have above-median-SES.

We construct standard errors for EC in each location using a bootstrap resampling method that adjusts for correlations in connectedness across individuals arising from having common pools of friends (see Supplementary Information B.3). Because sample sizes are large, almost none of the geographic difference in EC is due to sampling variation. At the county level, the mean standard error of 0.004 is more than an order of magnitude smaller than the signal standard deviation of EC across counties of 0.18. If we randomly split the microdata into two halves and estimate ECs by county in each half, we obtain a split-sample correlation (reliability) of 0.999 across counties, weighting by the number of people in each county with household income below the national median. The ZIP code-level estimates we release are also very precise, with a below-median-income-population-weighted split sample reliability of 0.99 (pooling all ZIP codes in the U.S.).

Childhood Economic Connectedness. We construct two measures of childhood EC: one based on links between individuals and their parents in our Facebook analysis sample and another based on data from Instagram.

To measure childhood EC in the Facebook sample, we restrict the sample to individuals whom we can link to high schools and to their parents (about 27% of the full sample). We assign parental SES ranks (estimated using the machine learning algorithm described in Section VI.B) within this subsample, ranking individuals based on parental SES relative to others in the same birth cohort. We then measure $f_{Q,i}$ as the fraction of friends from parental SES quantile Q within the subset of high school friends: friends who attended the same high school and are within three cohorts of the individual (so that they would have most likely overlapped in school). Ideally, we would directly observe all friendships made during childhood. However, because the Facebook platform was not available when the members of the birth cohorts we analyze were growing up, we use current friends who attended the same high school to identify friendships made in childhood. When calculating childhood EC by location, we assign individuals to the counties where their high schools are located rather than counties where they currently live, in order to map people to the places where they grew up. We do not produce ZIP code-level measures of childhood EC because we cannot reliably infer individuals’ childhood ZIP codes from their high schools’ locations (since children from many ZIP codes might attend a given school).

To measure childhood EC for users of Instagram, a widely-used social networking platform owned by Meta, we restrict the raw Instagram data to personal users (not business pages) in the United States who had not deactivated their account, been active on the platform within the last 30 days, and were predicted to be between 13 and 17 years of age as of May 28, 2022 (see Supplementary Information A.4 for further details). Next, we assign the individuals in our sample to ZIP codes based on their IP address and other features. Then, we assign Instagram users an SES estimate based on two variables: (i) the median household income of their residential ZIP code from publicly available data on incomes in the 25–44 age bin from the 2014–2018 American Community Survey, and (ii) the price of their phone. We then construct a weighted z-score of these two inputs, placing 2/3 weight on median household income and 1/3 weight on phone price. The higher weight on ZIP code-based income relative to phone price reflects that ZIP codes played a particularly large role in the machine learning model used to construct our baseline measures of SES in the Facebook data (though using other weights in the construction of the z-score produces similar results). We rank users nationally on the basis of these weighted z-scores to assign them an SES percentile rank. Users above the 50th percentile are termed high-SES; those at the 50th percentile and below are termed low-SES. Finally, we construct measures of individual economic connectedness as defined in equation 2. Because ties on Instagram, which are termed “follows,” are directional—that is, one person can follow another without that person following them—we restrict our attention to reciprocal followers in order to mimic friendships on Facebook when measuring connectedness.

Each of the two measures of childhood EC has certain advantages and drawbacks. The Facebook parental SES measure has the advantage of capturing the childhood friendships of individuals in approximately the same set of cohorts for which we measure economic mobility. However, because we are able to construct this measure only for the 27% of individuals whom we can link to parents and who report their high school, these estimates are noisier and potentially less representative than our baseline estimates. The Instagram data do not require parental linkage and capture all friends, not just high school friends, thereby yielding a larger and more comprehensive sample. The drawback of the Instagram EC measure is that it measures economic connectedness among the 2005–09 birth cohorts, rather than the 1978–83 cohorts for which we measure economic mobility. However, the stability of both economic mobility (Chetty et al. 2018) and economic connectedness (Supplementary Figure 1) within a location over time mitigates the consequences of this misalignment.

VI.E Measuring Cohesiveness

We represent a set of friendships by the matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$, where $A_{ij} = 1$ denotes the existence of a friendship (edge) between individuals i and j , and $A_{ij} = 0$ denotes the absence of a friendship. We focus on three measures of the structure of \mathbf{A} : clustering and support ratio, which are measures of local correlation in friendships, and spectral homophily, a measure of overall network fragmentation. Other measures of cohesiveness, such as algebraic connectivity Fiedler 1973, are also informative, but are difficult to compute or even approximate for networks of the scale we analyze. The three measures of cohesiveness we focus on here have the advantage of being computationally tractable in large samples.

Clustering. Coleman (1988) argued that if person i is friends with both persons j and k , then having j and k be friends with each other can help them collectively pressure and sanction person i , thereby helping to enforce norms and inducing pro-social behavior and investment. Motivated by this logic, many studies have measured the extent of such “network closure” by the degree of clustering within a person’s network: the frequency with which two friends of that person are in turn friends with each other. Letting $N_i(\mathbf{A})$ denote the set of i ’s friends and $d_i(\mathbf{A})$ its cardinality (the number of friends i has), the clustering of i ’s network is defined as

$$\text{Clustering}_i(\mathbf{A}) = \sum_{k, j \in N_i(\mathbf{A}), k < j} \frac{A_{kj}}{d_i(\mathbf{A})(d_i(\mathbf{A}) - 1)/2}, \quad (4)$$

We measure clustering in a community c as the average of equation 4 across people living in that community:

$$\text{Clustering}_c = \frac{\sum_{i \in c} \text{Clustering}_i(\mathbf{A})}{N_c}. \quad (5)$$

Support Ratio. Letting \mathbf{A}^c denote the subset of friendships between individuals who are both members of community c , we measure a community c 's support ratio as the overall frequency with which pairs of friends have at least one friend in common, focusing only on the people and friendships within that community:

$$\text{Support Ratio}_c = \frac{|\{(ij) : i, j \in c, A_{ij}^c = 1, [(A^c)^2]_{ij} > 0\}|}{|\{(ij) : i, j \in c, A_{ij}^c = 1\}|}. \quad (6)$$

Spectral Homophily. Spectral homophily measures the extent to which a network is fragmented into separate groups, and relates to the speed of information aggregation in a network. A wide variety of algorithms can detect subcommunities (Fortunato and Hric 2016), and spectral homophily provides a simple measure of how strongly a network splits into such subcommunities. Formally, spectral homophily is the second largest eigenvalue of the degree-normalized (row-stochasticised) adjacency matrix $\mathbf{A}_s^c \in [0, 1]^{n \times n}$. We measure spectral homophily in each county based on the set of friendships among individuals in our primary sample living in that county; friendship matrices are too sparse to estimate spectral homophily reliably at the ZIP code level. In the rare instances when there are fully isolated nodes within a county, we calculated spectral homophily on the largest connected component, which usually makes up the vast majority of users living in a county.

VI.F Measuring Civic Engagement

Volunteering rate. We start with the set of all Facebook Groups in the United States that are predicted to be about volunteering or activism based on their titles and do not have the privacy setting “secret” enabled. To further improve this classification, we manually review the 50 largest groups in the US as well as the largest group in each state, and remove the very small number of groups that are clearly misclassified. We then define the volunteering rate as the share of Facebook users in an area who are a member of at least one volunteering or activism group.

Civic organizations. We start with the set of all Facebook Pages in the United States that are categorized as “public good” pages based on the page title and category. We then remove pages that do not have a website linked, do not have a description on their Facebook page, or do not have an address listed. We then assign the page to a ZIP code and county based on its listed address, and calculate the density of civic organizations as the number of such pages per 1,000 Facebook users in the area.

VI.G Correlations

We weight all correlations and regressions by the number of individuals with below-national-median parental income as calculated using Census data (Chetty et al. 2018), unless otherwise noted. We cluster standard errors in county-level regressions by commuting zone and ZIP code-level regressions by county to adjust for potential spatial autocorrelation in errors, unless otherwise noted.

The causal effect estimates used in the Causal Effects of Place vs. Selection section are identified solely from individuals who move across areas and thus are much less precise than the baseline observational estimates of economic mobility used in the rest of the paper, making it necessary to adjust for attenuation bias in those correlation estimates due to sampling error. We adjust for attenuation bias by dividing the raw correlation between the causal estimates of mobility and economic connectedness by the square root of the reliability of the causal estimates of mobility, as estimated by Chetty and Hendren (2018a). The causal effect estimates are also unavailable at the ZIP code level due to small sample sizes for ZIP code-level moves. This is why we focus on the observational estimates of upward income mobility in our baseline analysis.

VI.H Privacy and Ethics

This 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 at 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 and Sample Construction

In this section, we provide further details on how we construct the data we use for our analysis, expanding on the discussion in Section VI.

A.1 Sample Coverage

Since much of our analysis relies on variation across areas, it is important that our sample has good coverage not just nationally but also across locations. Supplementary Figure 12 shows the geographic distribution of relative Facebook coverage rates, defined as the total number of Facebook users in our sample divided by population counts for that age group based on the 2014-2018 ACS, normalized by the national average coverage rate. Rates of coverage are high across the country, although usage rates are slightly lower in California. The correlations between sample sizes in the Facebook data and population counts are 0.99 across counties and 0.91 across ZIP codes.

Perhaps most important for our analysis, Facebook coverage does not vary systematically across locations with different income levels or demographic characteristics. Median household incomes, racial shares, and levels of education (based on ACS data) by county of residence are similar for Facebook users and the nationally representative ACS sample, as shown in Table 6. For example, median county incomes for individuals in our sample differ from those for the ACS by \$193 on average, a small difference relative to the standard deviation of median incomes across counties of more than \$15,000. Supplementary Figure 13 confirms that our coverage is consistent across the full distribution of ZIP code incomes.

To evaluate whether our main findings might be biased by differences in rates of Facebook coverage across areas, we replicate our key results restricting the sample to the top 25% of counties in terms of Facebook coverage. We find similar results in this subsample of counties (Supplementary Figure 14).

A.2 Linking Individuals to Parents

To construct measures of parental SES, we link individuals in our primary analysis sample (ages 25 to 44) to their parents. To do so, we start from the subsample of individuals in our primary analysis sample who self-report their parents. Since many individuals do not report their parents, we then use three methods sequentially to impute linkages for individuals who do not self-report parents: public user-generated wall posts that provide some indication of parental relationships; matching based on age and last names (using hashed strings to protect confidentiality);⁶ and familial relationships (e.g., inferring parents based on information provided by self-reported siblings).

We evaluate the accuracy of the imputed parental links by computing the false positive rate in the subsample of individuals who self-report their parents. For such individuals, we assign mothers and fathers using the self-reports and imputation procedures separately. The match rates between the self-reports and the imputed linkages are 78% and 83% for mothers and fathers, respectively. The correlations of parental SES between the self-reported and imputed parental matches are 0.87 and 0.91 for mothers and fathers, respectively. Out of all final parent-child linkages, 39.4% are based on self-reports and the rest are based on imputations.

Overall, we match 46% of the individuals in our primary sample to parents. 31% of the primary sample is assigned a parental SES. We are unable to assign an SES to all parents because some of them are above age 65 in 2022 and we do not assign SES to those over age 65 since we cannot measure SES reliably for retired individuals.

⁶Hashing is a procedure that recodes a specific string into a unique numerical identifier. This enables us to perform name-matching across users while only using their corresponding numerical identifiers for privacy protection.

A.3 Identifying 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. We also exclude self-reported schools where users have fewer than 10 friends (with ages within three years of their own). 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.

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.

A.4 Instagram Data

We construct the Instagram sample used to measure childhood EC by restricting to personal users (not business pages) in the United States who had not deactivated their account, had been active on the platform within the last 30 days, and were predicted to be between 13 and 17 years of age as of May 28, 2022. We obtain this age range prediction from a Meta-internal machine learning model, which is trained on multiple signals including Instagram accounts linked to Facebook accounts (see <https://about.fb.com/news/2021/07/age-verification/> for details on this model). We remove influencers and prominent public figures from the sample, defined as accounts that are above the 99th percentile in terms of followers or that receive more than 80,000 messages per week. Lastly, we remove accounts that do not have at least 5 reciprocal follows; that is, we require that at least five accounts a user follows must follow them back.

Supplementary Figure 15 plots the mean SES percentile rank of individuals’ friends against their own SES percentile rank, replicating Figure 1 in the Instagram sample. We find slightly greater homophily in the Instagram sample than in our baseline analysis sample, with an SES rank-rank slope of 0.51 (and a slope of 0.47 between the 10th and 90th percentiles of own SES).

A.5 External 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.5.1 Neighborhood Characteristics

American Community Survey. We obtain data on median incomes by Census block group and ZIP code from the 2014–2018 American Community Survey (ACS). These block-group-level income data are used in our machine learning algorithm for predicting socioeconomic status (see Supplementary Information B.1).

We also use the ACS to construct measures of racial and income segregation across tracts within each county. We measure racial segregation using Theil’s H index, following equation (4) in Chetty et al. (2014). We compute the racial segregation index based on the shares of four groups in each tract: whites, Blacks, Hispanics, and all others. We measure income segregation using the generalized H index averaging across all income percentiles introduced by Reardon and Bischoff (2011), following equation (5) from Chetty et al. (2014). We are unable to construct reliable estimates of the segregation variables for counties with populations below 20,000 and cannot construct estimates for counties that consist of a single tract. We therefore omit those counties when analyzing income and racial segregation.

Opportunity Atlas. 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.

We also use the following variables from the Opportunity Atlas and Chetty et al. (2014), which are derived from the ACS and other sources, for correlational analyses: jobs within 5 miles, job growth rate 2004–2013, employment rate in 2000, Gini coefficient, top 1% share, share above poverty line, mean household income, mean 3rd grade math score, share college graduates, Black share, Hispanic share, and single parent share. We measure income inequality in each county as the raw Gini coefficient estimated using tax data minus the income share of the top 1% to obtain a measure of inequality among the bottom 99%, which Chetty et al. (2014) show is most predictive of differences in upward mobility across areas. We exclude four small counties where the resulting estimate of the Gini coefficient is negative.

A.5.2 Add Health Survey Data

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of students who attended grades 7–12 in 1994–1995. We use data from the first wave of Add Health (1995), which contains information on students’ self-reported friendship networks as well as household income reported by the female head of household (unless there is no such individual in the household). The sample consists of 18,924 students, 90 percent of whom have information regarding parental household income. The social network data in the Add Health survey was constructed by asking the teenagers to nominate at most 5 male and 5 female friends. We define an individual’s set of friends as those they nominate as one of their friends as well as all individuals who list them as one of their friends. Figure 1b is generated by ranking students according to their household income and averaging friends’ household income ranks for each student.

A.5.3 External Social Capital Measures

Penn State Index. We use the county-level social capital index constructed by Rupasingha et al. (2006) as an index of civic engagement. This index is the first principal component of a set of four variables: the Census response rate; voter turnout in presidential elections; the number of per capita non-profit organizations; and an aggregate of a set of variables containing the number of various organizations (such as religious, civic, and labor organizations).

Local Trust Index. We measure local trust using the state-level social support subindex from Social Capital Project (2018). It comprises four indicators derived from survey data: the share of adults who “get the social and emotional support [they] need”; the share of adults who do favors for neighbors at least once a month; the share who trust most or all of their neighbors; and the average number of “close” friends reported.

B Supplementary Methods

This section provides further details on three aspects of our methods: (1) our algorithm for measuring socioeconomic status; (2) the construction of certain social capital measures we analyze in Supplementary Tables; and (3) estimation of standard errors for our publicly available estimates of economic connectedness.

B.1 Measuring Socioeconomic Status

We construct our baseline measure of socioeconomic status by combining various proxies for SES (e.g., median incomes in one’s residential ZIP code, cell phone model, college attended, etc.) that are observed in the Facebook data. We estimate SES solely for the purpose of this research project, and delete these SES measures at the end of our analysis; the SES measure we use is not an internal measure provided by Facebook.

The underlying proxies we use for SES, which are listed in Supplementary Table 5, can be combined in many ways to create a single SES measure. We seek to identify the combination of proxies that best predicts household median income at the block group level using a machine learning model. To identify this combination, we begin by forming a training sample that consists of users aged 25 to 64 in our primary analysis sample with Location History (LH) enabled. We observe the residential Census block group of LH users, and use this information to assign each of them the median household income in their residential block group using data from the publicly available 2014–2018 American Community Survey (ACS), separately for the 25–44 and 45–64 year age buckets.

Next, we train a gradient-boosted regression tree to predict the log of block group median income in the LH sample, using the SES proxies described in Supplementary Table 5 as predictors. Because these predictors are available for non-LH users as well, this model allows us to create SES predictions for all users in our analysis sample. To reduce the risk of overfitting, we impose a maximum tree depth and hold out 10% of our data as a validation sample. Reassuringly, we find that the model has similar performance in the validation sample.

After estimating the model, we construct predicted values of SES for the entire sample (including non-LH users). We then convert the individual-level SES predictions obtained from the model to local rankings of individuals within each county. We map these local rankings to ranks in the national SES distribution by using data on income distributions by county from the ACS, which releases the number of households in 16 income bins in each county. For each county, we fit a parametric distribution to this discrete data, using Stata’s `mgbe` (multimodel generalized beta estimator) command, as described in von Hippel et al. (2015). The distribution is selected using maximum likelihood estimation with the following log-likelihood function:

$$l(\theta|X) = \ln \prod_{b=1}^B (P(l_b < X < u_b))^{n_b} = \ln \prod_{b=1}^B (F(u_b) - F(l_b))^{n_b}, \quad (7)$$

where for each income bin $b \in \{1, \dots, B = 16\}$, the lower and upper boundaries are denoted by l_b and u_b , and the number of observations in the bin by n_b . We estimate the parameters, θ , for the cumulative distribution function, $F(\cdot)$, that maximize the likelihood of each observation falling in its respective income bin (i.e. $P(l_b < X < u_b)$). We perform a similar exercise for each of the seven distributions in the generalized beta family (dagum, singh-maddala, beta, loglogistic, gamma, generalized gamma, and weibull), and choose the distribution with the smallest Akaike information criterion (AIC). Using the estimated local income distribution, we translate the local rankings in each county obtained from the machine learning model to estimates of levels of income. Lastly, we rank all individuals in the national distribution relative to others in their birth cohort to obtain our final SES ranks.

For sensitivity analyses, we also construct three other measures of socioeconomic status:

ACS Block Group Median Income. We assign each user in the LH sample the household median income for their residential block group as described above.

ACS ZIP Code Household Median Income. To construct an area-level proxy that is available for the entire primary analysis sample, we assign each user the household median income for their ZIP code based on the 2014–18 ACS. ZIP code is available for all users, not just those in the LH sample.

Z-Score Index. We create a different combination of the SES proxies that does not rely on the machine learning model by taking a mean of the z-scores of the following six variables: the ACS ZIP code median household income, days since joining Facebook, phone price, college tier, web usage, and an indicator for self-reported graduate school. We focus on this set of variables because they are most predictive of block-group SES in the ML model and have no missing observations. We construct the z-score for each variable by subtracting the mean and dividing by the standard deviation, separately by birth cohort. The composite index is then constructed as the unweighted average of the z-scores, with each variable signed so that higher values correlate with higher SES predictions from the machine learning model.

As above, we create SES ranks based on all of these measures by ranking individuals relative to all others in the primary sample in their birth cohort.

Supplementary Table 6 shows how measures of economic connectedness based on each of these SES proxies correlate with each other and with economic mobility across counties. All the measures are highly correlated with each other and with upward income mobility, indicating that our results are insensitive to the specific algorithm used to measure SES.

B.2 Other Social Capital Measures

In this section, we describe other social capital measures that we analyze in the Supplementary Information but do not discuss in the main text.

Mean Friend Rank for Individuals at p25. To construct this measure, we first calculate the mean SES rank of each individual’s friends. In each area (county or ZIP), we then regress this mean friend rank on the individual’s own rank and calculate the predicted average friend rank from this regression for individuals at the 25th percentile of the national SES distribution. The difference between this measure and our baseline measure of economic connectedness is that it controls for differences in the SES distribution among individuals with below-median SES across areas.

Bottom-to-Top SES Quintile EC. Analogous to our baseline economic connectedness measure, this statistic is five times the share of top-SES-quintile friends among bottom-SES-quintile individuals.

Spectral SES Homophily. This is a summary measure of the degree of connection across individuals in different quintiles of the national SES distribution in a given area c , as defined in Golub and Jackson (2012). We begin by constructing a five-by-five matrix whose i, j elements are the share of friends from quintile j among individuals in quintile i of the national SES distribution. We then define spectral SES homophily as the second eigenvalue in magnitude of this matrix, which ranges from 0 to 1, with higher values corresponding to more homophily by SES. If the matrix had 0.2 in every entry, then each quintile would have 1/5 of its friendships in every quintile, and there would be no homophily. The second eigenvalue of such a matrix is 0—corresponding to 0 homophily by SES. If the community were fully homophilistic and had no connections across SES quintiles, then the matrix would be the identity matrix and would have a second eigenvalue of 1.

B.3 Standard Errors of Economic Connectedness Estimates

This section explains how we construct standard errors for our estimates of economic connectedness for each county and ZIP code in our sample (as well as for high schools and colleges in our companion paper, Chetty et al. 2022).

Recall that we define EC in a given area c as the mean level of individual economic connectedness (IEC) of low-SES (below-median) users:

$$\text{EC}_c = \frac{\sum_{i \in L \cap c} \text{IEC}_i}{N_{Lc}}, \quad (8)$$

where N_{Lc} is the number of low-SES users in c .

Since this is a mean of individual-level values, a natural estimator for the standard error of the EC estimate that ignores the network structure of the data is:

$$\text{se}_{\text{naive}}(\text{EC}_c) = \sqrt{\frac{\sum_{i \in L \cap c} (\text{IEC}_i - \text{EC}_c)^2}{N_{Lc} - 1}} \quad (9)$$

However, this “naive” standard error estimate—which assumes that IEC_i is independent and identically distributed across users within c —is likely to be an underestimate of the true standard error of our EC estimate because IEC is likely to be correlated across people in a given community c . Perhaps the most important source of such correlation is that all users in a given community draw their friends from the same pool of potential friends (e.g., the set of students who attend a given school), which is itself stochastic given the limited size of each group.

We correct for this additional source of variance in our estimates using the bootstrap-based approach described in Section 2.3 of Davezies et al. (2021). Each iteration of the bootstrap yields a potential network that could have been observed, and hence a corresponding value of EC that could have been observed. We calculate the standard error of EC in a given cell c as the standard deviation of EC across these potential realizations. Each iteration of the bootstrap consists of the following steps:

1. Assign each individual i in the analysis sample a weight π_i from a Poisson(1) distribution, reflecting the number of times the user is “sampled” in this bootstrap iteration.
2. Construct a sampled friend list from friendships in the original Facebook graph. For two users i and j with Poisson weights π_i and π_j who are friends, their friendship appears $\pi_i \times \pi_j$ times in the sampled friend list. Note that if $\pi_j = 0$, this friendship will not appear in the sampled friend list.
3. Calculate a new IEC for each individual i using this sampled friend list. For a given individual i we calculate their new IEC as two times the weighted (using weights π_j from step 1, where j indexes friends of i) proportion of their friends with high SES.
4. Take the weighted average of IEC over individuals i in the cell using the weights π_i from step 1.

The average standard error of EC in each setting using this bootstrap method is similar to an estimate of the standard deviation of the noise component of EC across groups within each setting based on split-sample estimates of the reliability of the EC estimates. We estimate reliability by randomly splitting the nodes in each group into two and then correlate EC estimates in the two split samples with each other. We then use this reliability estimate to calculate the portion of the total variance in EC that is due to noise within each setting.

Note that there may be other sources of correlation between the IECs of individuals in a given cell c that are not fully captured by the bootstrap method or split sample reliability calculation, so the standard error estimates we report should be interpreted with caution.

C Supplementary Discussion

C.1 Prior Work with Facebook Data

Facebook data have been used to study the effects of social networks on a variety of outcomes in prior work: patent citations (Bailey et al. 2018a), home purchasing decisions (Bailey et al. 2018b), mortgage choices (Bailey et al. 2019a), cell phone adoption (Bailey et al. 2019b), labor market outcomes (Gee 2018; Gee et al. 2017), commuting flows (Bailey et al. 2020a), international trade flows (Bailey et al. 2021), investment decisions (Kuchler et al. 2022a), peer-to-peer lending (Allen et al. 2020), EITC claiming behavior (Wilson 2020), racial homophily (Wimmer and Lewis 2010), health behavior and beliefs (Bailey et al. 2020b), mortality rates (Hobbs et al. 2016), the spread of COVID-19 (Kuchler et al. 2022b), and the social integration of international migrants (Bailey et al. 2022). Although they use some of the same underlying data, prior studies have not constructed systematic measures of social capital or studied their determinants as we do here.

C.2 Associations Between Civic Engagement and Economic Mobility

Our results on the lack of a strong association between civic engagement and economic mobility may appear to be inconsistent with the findings of the Social Capital Project, who report stronger correlations between measures of what we term civic engagement and economic mobility (Social Capital Project 2018). The difference is explained primarily by the fact that we weight our correlations by the number of children with below-national-median parental income (which are very similar to population weights), whereas Social Capital Project reports unweighted correlations (Social Capital Project 2018). The unweighted state-level correlation between the Social Capital Project “social support” measure and our upward mobility measure is 0.50, whereas the weighted correlation is 0.17 (in contrast, weighted and unweighted correlations of economic connectedness with upward mobility are very similar). This is because civic engagement is more highly correlated with mobility (as well as connectedness) in rural areas. A further difference is that we measure economic mobility here as the mean adult income rank of individuals with low-income (25th percentile) parents, whereas the Social Capital Project focuses on a relative mobility measure—the difference between outcomes at the 25th and 75th percentile (Social Capital Project 2018). Economic connectedness remains one of the strongest predictors of relative mobility as well.

C.3 Associations Between Social Capital and Life Expectancy

We measure the association between social capital and life expectancy using publicly available data on life expectancy by county for individuals at the 25th percentile of the national income distribution from Chetty et al. (2016c). Clustering coefficients and support ratios are much stronger predictors than economic connectedness of differences in life expectancy among low-income individuals across counties (Supplementary Figure 11). Perhaps surprisingly, higher levels of clustering and support are associated with *lower* life expectancy. One potential explanation for this correlation is that areas with relatively low life expectancy face challenges that create a demand for closely-knit communities to provide support.

C.4 Friendship Rates by SES Percentile Rank

In Supplementary Figure 16a, we plot friendship rates by own and friends’ SES percentile ranks, showing the fraction of friends from each SES rank by an individual’s own SES percentile rank. Individuals with similar SES are more likely to be friends with each other. Consistent with the higher slope in the upper tail of Figure 1, individuals in the upper tail of the SES distribution are especially likely to befriend those in the upper tail (Supplementary Figure 16b). 9% of the friends of individuals in the top 1% of the national SES distribution come from the top 1%; hence,

individuals in the top 1% are 9 times more likely to befriend those in the top 1% than would occur in the absence of homophily.

References

- Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna, and Benjamin A Olken (2016). “Network structure and the aggregation of information: Theory and evidence from Indonesia”. *American Economic Review* 106.7, pp. 1663–1704.
- Allen, Linda, Lin Peng, and Yu Shan (2020). “Social Networks and Credit Allocation on FinTech Lending Platforms”. Working Paper.
- Ambrus, Attila, Markus Mobius, and Adam Szeidl (2014). “Consumption risk-sharing in social networks”. *American Economic Review* 104.1, pp. 149–82.
- Aral, Sinan and Christos Nicolaides (2017). “Exercise Contagion in a Global Social Network”. *Nature Communications* 8.1, p. 14753.
- 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 (2019a). “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 Richmond, and Johannes Stroebel (2021). “International Trade and Social Connectedness”. *Journal of International Economics* 129, p. 103418.
- Bailey, Michael, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2020b). *Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing during the COVID-19 Pandemic*. Working Paper 28234. National Bureau of Economic Research.
- Bailey, Michael, Drew M Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2022). *The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany*. Working Paper 29925. National Bureau of Economic Research.
- Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2019b). *Peer Effects in Product Adoption*. Working Paper 25843. National Bureau of Economic Research.
- Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006). “Who’s who in networks. Wanted: The key player”. *Econometrica* 74.5, pp. 1403–1417.
- Banfield, Edward C. (1958). *The moral basis of a backward society*. Free Press: New York.
- Beaman, Lori A. (2012). “Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.” *The Review of Economic Studies* 79.1, pp. 128–161.
- Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer (2019). *Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice*. Working Paper 26164. National Bureau of Economic Research.
- Blau, Peter M. (1977). “A Macrosociological Theory of Social Structure”. *American Journal of Sociology* 12.1, pp. 26–54.
- Bolte, Lukas, Nicole Immorlica, and Matthew O. Jackson (2020). “The Role of Referrals in Inequality, Immobility, and Inefficiency in Labor Markets”. *SSRN*: <https://ssrn.com/abstract=3512293>.
- Bourdieu, Pierre (1986). “The Forms of Capital”. *Handbook of Theory and Research for the Sociology of Education*. Ed. by John Richardson. Westport, Conn: Greenwood, pp. 15–29.
- Brooks, David (2020). “Who Is Driving Inequality? You Are”. *The New York Times*.
- Burchardi, Konrad B. and Tarek A. Hassan (2013). “The Economic Impact of Social Ties: Evidence from German Reunification”. *The Quarterly Journal of Economics* 128.3, pp. 1219–1271.

- Calvo-Armengol, Antoni and Matthew O. Jackson (2004). “The effects of social networks on employment and inequality”. *The American Economic Review* 94.3, pp. 426–454.
- Carrell, Scott E., Mark Hoekstra, and James E. West (2011). “Is Poor Fitness Contagious? Evidence from Randomly Assigned Friends”. *Journal of Public Economics* 95.7, pp. 657–663.
- Centola, Damon (2018). *How Behavior Spreads: The Science of Complex Contagions*. Vol. 3. Princeton University Press.
- Centola, Damon, Víctor M. Eguíluz, and Michael W. Macy (2007). “Cascade dynamics of complex propagation”. *Physica A: Statistical Mechanics and its Applications* 374.1, pp. 449–456.
- 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, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang (2017). “The Fading American Dream: Trends in Absolute Income Mobility since 1940”. *Science*. eprint: <http://science.sciencemag.org/content/early/2017/04/21/science.aal4617.full.pdf>.
- Chetty, Raj and Nathaniel Hendren (2018a). “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects”. *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.
- (2018b). “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates”. *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.
- Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2020). “Race and Economic Opportunity in the United States: An Intergenerational Perspective”. *The Quarterly Journal of Economics* 135.2, pp. 711–783.
- Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz (2016a). “The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment”. *American Economic Review* 106.4, pp. 855–902.
- (2016b). “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment”. *American Economic Review* 106.4, pp. 855–902.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.
- Chetty, Raj, M Stepner, S Abraham, S Lin, B Scuderi, N Turner, A Bergeron, and D Cutler (2016c). “The Association between Income and Life Expectancy in the United States, 2001–2014”. *Journal of the American Medical Association* 315.16, pp. 1750–1766.
- Chetty, Raj et al. (2022). “Social Capital II: Determinants of Economic Connectedness”. *NBER Working Paper*.
- Chyn, Eric (2018). “Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children”. *American Economic Review* 108.10, pp. 3028–56.
- Coleman, James S. (1988). “Social Capital in the Creation of Human Capital”. *American Journal of Sociology* 94, S95–S120.
- Corak, Miles (2013). “Income Inequality, Equality of Opportunity, and Intergenerational Mobility”. *Journal of Economic Perspectives* 27.3, pp. 79–102.
- Cutler, David M. and Edward L. Glaeser (1997). “Are Ghettos Good or Bad?” *The Quarterly Journal of Economics* 112.3, pp. 827–72.
- Davezies, Laurent, Xavier D’Haultfoeuille, and Yannick Guyonvarch (2021). “Empirical Process Results for Exchangeable Arrays”. *The Annals of Statistics* 49.2, pp. 845–862.
- DeFilippis, James (2001). “The myth of social capital in community development”. *Housing Policy Debate* 12.4, pp. 781–806. eprint: <https://doi.org/10.1080/10511482.2001.9521429>.
- Deutscher, Nathan (2020). “Place, peers, and the teenage years: long-run neighborhood effects in Australia”. *American Economic Journal: Applied Economics* 12.2, pp. 220–49.

- 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.
- Durlauf, Steven N. and Ananth Seshadri (2017). “Understanding the Great Gatsby Curve”. Eichenbaum, Martin and Jonathan A. Parker. *NBER Macroeconomics Annual 2017, volume 32*. University of Chicago Press, pp. 333–393.
- Eagle, Nathan, Michael Macy, and Rob Claxton (2010). “Network Diversity and Economic Development.” *Science* 328.5981.
- Fiedler, Miroslav (1973). “Algebraic connectivity of graphs”. *Czechoslovak Mathematical Journal* 23.2, pp. 298–305.
- Fortunato, Santo and Darko Hric (2016). “Community detection in networks: A user guide”. *Physics Reports* 659, pp. 1–44.
- 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.
- Glaeser, Edward L., David I. Laibson, José A. Scheinkman, and Christine L. Soutter (2000). “Measuring Trust*.”. *The Quarterly Journal of Economics* 115.3, pp. 811–846. eprint: <https://academic.oup.com/qje/article-pdf/115/3/811/5174945/115-3-811.pdf>.
- Golub, Benjamin and Matthew O. Jackson (2012). “How Homophily Affects the Speed of Learning and Best-Response Dynamics”. *Quarterly Journal of Economics* 127.3, pp. 1287–1338.
- Hedefalk, Finn and Martin Dribe (2020). “The social context of nearest neighbors shapes educational attainment regardless of class origin”. *Proceedings of the National Academy of Sciences* 117.26, pp. 14918–14925. eprint: <https://www.pnas.org/content/117/26/14918.full.pdf>.
- Herdağdelen, Amaç, Lada Adamic, and Bogdan State (2021). “Correlates of Social Capital In Facebook Groups”.
- Hill, Allison L., David G. Rand, Martin A. Nowak, and Nicholas A. Christakis (2010). “Emotions as infectious diseases in a large social network: the SISa model.” *Proc. R. Soc. B.* 277.1701, pp. 3827–35.
- Hobbs, William R., Moira Burke, Nicholas A. Christakis, and James H. Fowler (2016). “Online social integration is associated with reduced mortality risk.” *Proceedings of the National Academy of Sciences* 113.46, pp. 12980–12984.
- Jackson, Matthew O. (2019). *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. Pantheon Books: New York.
- (2020). “A Typology of Social Capital and Associated Network Measures”. *Social Choice and Welfare* 54.2, pp. 311–336.
- (2021). “Inequality’s Economic and Social Roots: the Role of Social Networks and Homophily”. *SSRN*: <https://ssrn.com/abstract=3795626>.
- Jackson, Matthew O., Tomas R. Rodriguez-Barraquer, and Xu Tan (2012). “Social Capital and Social Quilts: Network Patterns of Favor Exchange”. *American Economic Review* 102.5, pp. 1857–1897.
- Jackson, Matthew O. and Evan C. Storms (2018). “Behavioral Communities and the Atomic Structure of Networks”. *Arxiv*: <https://arxiv.org/abs/1710.04656>.
- 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.
- Knack, Stephen and Philip Keefer (1997). “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation”. *The Quarterly Journal of Economics* 112.4, pp. 1251–1288.
- Krueger, Alan (2012). *The Rise and Consequences of Inequality in the United States*. Tech. rep. Center for American Progress.

- Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou (2022a). “Social Proximity to Capital: Implications for Investors and Firms”. *The Review of Financial Studies* 35.6, pp. 2743–2789.
- Kuchler, Theresa, Dominic Russel, and Johannes Stroebel (2022b). “JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook”. *Journal of Urban Economics* 127, p. 103314.
- Kuchler, Theresa and Johannes Stroebel (2022). *Social Interactions, Resilience, and Access to Economic Opportunity: A Research Agenda for the Field of Computational Social Science*. CESifo Working Paper 9606.
- Lancee, Bram (2010). “The economic returns of immigrants’ bonding and bridging social capital: The case of the Netherlands”. *International Migration Review* 44.1, pp. 202–226.
- Laschever, Ron (2013). *The Doughboys Network: Social Interactions and the Employment of World War I Veterans*. SSRN Scholarly Paper ID 1205543. Rochester, NY: Social Science Research Network.
- 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>.
- Lin, Nan (1999). “Building a network theory of social capital”. *Connections* 22.1, pp. 28–51.
- Lin, Nan and Mary Dumin (1986). “Access to occupations through social ties”. *Social Networks* 8.4, pp. 365–385.
- List, John A, Fatemeh Momeni, and Yves Zenou (2020). *The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods*. Working Paper 28283. National Bureau of Economic Research.
- Loury, Glenn C. (1977). “A Dynamic Theory of Racial Income Differences”. *Women, Minorities, and Employment Discrimination*. Ed. by P.A. Wallace and Annette LaMond. Lexington, MA: Lexington Books, pp. 133–186.
- Maas, Paige, Shankar Iyer, Andreas Gros, Wonhee Park, Laura McGorman, Chaya Nayak, and P. Alex Dow (2019). “Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery”. *Proceedings of the 16th ISCRAM Conference*.
- Manduca, Robert and Robert J. Sampson (2019). “Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children”. *Proceedings of the National Academy of Sciences* 116.16, pp. 7772–7777. eprint: <https://www.pnas.org/content/116/16/7772.full.pdf>.
- Montgomery, James D. (1991). “Social networks and labor-market outcomes: Toward an economic analysis”. *The American Economic Review* 81.5, pp. 1408–1418.
- Nannicini, Tommaso, Andrea Stella, Guido Tabellini, and Ugo Troiano (2013). “Social capital and political accountability”. *American Economic Journal: Economic Policy* 5.2, pp. 222–250.
- Perrin, Andrew and Monica Anderson (2019). “Share of U.S. Adults Using Social Media, Including Facebook, Is Mostly Unchanged since 2018”. *Pew Research Center: Fact Tank*.
- Portes, Alejandro (1998). “Social capital: Its origins and applications in modern sociology”. *Annual Review of Sociology* 24.1, pp. 1–24.
- Putnam, Robert (1995). “Bowling Alone: America’s Declining Social Capital”. *Journal of Democracy* 6.1, pp. 65–78.
- (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster.
- (2016). *Our Kids: The American Dream in Crisis*. Simon and Schuster.
- Putnam, Robert, Robert Leonardi, and Raffaella Y Nanetti (1994). *Making democracy work*. Princeton University Press.
- Rand, David G., Samuel Arbesman, and Nicholas A. Christakis (2011). “Dynamic social networks promote cooperation in experiments with humans.” *Proceedings of the National Academy of Sciences* 108.48, pp. 19193–19198.

- Rand, David G., Martin A. Nowak, James H. Fowler, and Nicholas A. Christakis (2014). “Static network structure can stabilize human cooperation.” *Proceedings of the National Academy of Sciences* 111.48, pp. 17093–17098.
- Reardon, S and Kendra Bischoff (2011). “Income Inequality and Income Segregation”. *American Journal of Sociology* 116.4, pp. 1092–1153.
- 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.
- Rupasingha, Anil, Stephan J Goetz, and David Freshwater (2006). “The Production of Social Capital in US Counties”. *The Journal of Socio-Economics* 35.1, pp. 83–101.
- 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.
- Simmel, Georg (1902). “The metropolis and mental life”. *The urban sociology reader*. Routledge, pp. 37–45.
- Small, Mario Luis (2010). *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Reprint edition. Oxford University Press.
- Social Capital Project (2018). *The Geography of Social Capital in America*. Tech. rep. SCP Report Number 1-18.
- (2019). *The Wealth of Relations*. Tech. rep. SCP Report Number 3-19.
- Stack, Carol B. (1974). *All Our Kin: Strategies for Survival in a Black Community*. Harper and Row.
- Tabellini, Guido (2010). “Culture and institutions: economic development in the regions of Europe”. *Journal of the European Economic Association* 8.4, pp. 677–716.
- Thomas, William Isaac and Florian Znaniecki (1919). *The Polish peasant in Europe and America: Monograph of an immigrant group*. Vol. 3. University of Chicago Press.
- von Hippel, Paul T., Samuel V. Scarpino, and Igor Holas (2015). “Robust Estimation of Inequality from Binned Incomes”. *Sociological Methodology* 46, pp. 212–251.
- 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.
- Watts, Duncan J. (2004). *Six degrees: The science of a connected age*. WW Norton and Company.
- White, Karl R. (1982). “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.
- Wimmer, Andreas and Kevin Lewis (2010). “Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook.” *American Journal of Sociology* 116.2, pp. 583–642.

TABLE 1: Conditional Distributions of Friends' SES by Own SES Decile

A. Proportion of Friends by Own SES Decile										
Own Decile	Proportion of Friends in Decile									
	1	2	3	4	5	6	7	8	9	10
1	23.4	17.5	13.7	11.0	9.2	7.7	6.2	4.9	3.8	2.5
2	15.8	15.9	13.8	11.9	10.6	9.2	7.9	6.5	5.1	3.4
3	11.7	12.8	12.8	12.1	11.4	10.5	9.4	8.1	6.7	4.6
4	9.0	10.5	11.4	11.7	11.7	11.4	10.6	9.6	8.2	5.9
5	7.0	8.7	10.0	10.8	11.6	11.9	11.6	11.0	9.8	7.5
6	5.5	7.1	8.5	9.7	11.0	12.0	12.5	12.4	11.7	9.6
7	4.3	5.7	7.1	8.5	10.0	11.7	13.0	13.7	13.8	12.2
8	3.3	4.6	5.9	7.2	8.9	11.0	12.9	14.7	15.9	15.6
9	2.4	3.4	4.6	5.8	7.5	9.7	12.1	14.9	18.3	21.2
10	1.6	2.1	2.9	3.9	5.3	7.4	10.0	13.4	19.4	34.0

B. Number of Friends by Own SES Decile										
Own Decile	Average Friends in Decile									
	1	2	3	4	5	6	7	8	9	10
1	110.6	80.3	65.7	52.8	43.5	36.0	28.9	22.6	17.0	11.1
2	80.3	76.4	68.7	59.0	51.3	44.4	37.0	30.1	23.3	15.0
3	65.7	68.7	68.4	62.7	57.3	51.9	45.2	38.3	30.8	20.5
4	52.8	59.0	62.7	61.9	59.8	57.0	52.1	46.1	38.7	27.0
5	43.5	51.3	57.3	59.8	62.1	62.8	60.4	56.2	49.5	36.5
6	36.0	44.4	51.9	57.0	62.8	68.1	69.7	68.6	64.1	50.8
7	28.9	37.0	45.2	52.1	60.4	69.7	76.7	80.5	80.4	69.5
8	22.6	30.1	38.3	46.1	56.2	68.6	80.5	91.3	98.7	94.4
9	17.0	23.3	30.8	38.7	49.5	64.1	80.4	98.7	120.5	136.8
10	11.1	15.0	20.5	27.0	36.5	50.8	69.5	94.4	136.8	240.0

C. Proportion of High-School Friends by Parental SES Decile										
Own Parental Decile	Proportion of Friends in Decile									
	1	2	3	4	5	6	7	8	9	10
1	17.1	13.4	12.0	10.8	9.9	9.1	8.3	7.5	6.6	5.4
2	12.3	13.2	11.6	11.0	10.4	9.8	9.1	8.5	7.6	6.4
3	10.4	11.0	11.9	10.9	10.6	10.2	9.8	9.3	8.5	7.4
4	9.0	9.9	10.5	11.5	10.7	10.5	10.2	9.9	9.4	8.4
5	8.0	9.1	9.8	10.4	11.4	10.7	10.6	10.5	10.1	9.4
6	7.2	8.4	9.3	10.0	10.5	11.6	11.0	11.0	10.9	10.3
7	6.5	7.7	8.7	9.5	10.1	10.8	12.0	11.6	11.7	11.5
8	5.8	7.0	8.1	9.1	9.8	10.6	11.4	12.8	12.6	12.9
9	5.0	6.3	7.4	8.4	9.4	10.3	11.3	12.4	14.5	14.9
10	4.2	5.3	6.4	7.5	8.6	9.7	11.0	12.5	14.8	19.9

D. Number of High-School Friends by Parental SES Decile										
Own Parental Decile	Average Friends in Decile									
	1	2	3	4	5	6	7	8	9	10
1	9.9	8.3	7.8	7.3	6.9	6.5	6.0	5.6	4.9	4.1
2	8.0	8.5	8.2	8.0	7.8	7.6	7.2	6.8	6.2	5.2
3	7.2	8.0	8.7	8.6	8.6	8.5	8.2	7.9	7.3	6.4
4	6.6	7.6	8.4	9.1	9.1	9.2	9.0	8.9	8.4	7.5
5	6.2	7.4	8.3	9.0	9.7	9.7	9.7	9.8	9.5	8.7
6	5.8	7.1	8.1	9.0	9.6	10.4	10.4	10.6	10.5	9.9
7	5.3	6.7	7.8	8.8	9.6	10.3	11.2	11.4	11.6	11.3
8	4.9	6.2	7.5	8.6	9.5	10.4	11.3	12.4	12.7	12.9
9	4.3	5.6	6.9	8.1	9.2	10.3	11.5	12.7	14.4	15.2
10	3.6	4.8	6.0	7.2	8.4	9.7	11.2	12.9	15.2	20.2

Notes: This table shows the number and percentage of friends from each SES decile by individuals' own SES deciles. Panels A and B, respectively, calculate friend shares and average numbers of friends using individuals' own SES and entire friendship networks. Panels C and D calculate friend shares and average numbers of friends using parental SES and individuals' high school friendship networks—the set of peers within three birth cohorts who attended the same high school. In Panels A and B, the statistics are calculated on our primary analysis sample (see Table 6a). In Panels C and D, they are calculated on the subsample with linked parental SES and high schools (see Table 6b). See Supplementary Figure 16 for a heat map of an analogous matrix.

TABLE 2: Correlation Matrix for Social Capital Measures across Counties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Economic Connectedness (EC)	1.00								
(2) Language Connectedness	0.10	1.00							
(3) Age Connectedness	-0.45	0.17	1.00						
(4) Clustering	0.01	0.38	0.51	1.00					
(5) Support Ratio	-0.25	0.30	0.50	0.64	1.00				
(6) Spectral Homophily	-0.09	-0.37	-0.49	-0.61	-0.51	1.00			
(7) Penn State Index	0.31	0.08	-0.04	0.39	0.28	-0.25	1.00		
(8) Civic Organizations	0.27	0.16	0.05	0.37	0.23	-0.33	0.67	1.00	
(9) Volunteering Rate	0.46	0.28	-0.04	0.30	0.23	-0.35	0.44	0.46	1.00

Notes: This table reports county-level pairwise correlations of the primary social capital measures that we analyze, weighted by the number of children with below-median parental income in each county as calculated in the Opportunity Atlas (Chetty et al. 2018) using Census data. Economic connectedness is twice the share of above-median-SES friends among below-median-SES people. Language connectedness is the share of friends who set their Facebook language to English among users who do not set their language to English, divided by the national share of users who set their language to English. Age connectedness is the share of friends who are 35 to 44 among users who are 25 to 34, divided by the national share of users aged 35-44. Clustering is the fraction of an individual’s friend pairs who are also friends with each other, averaged over all individuals in the county. Support ratio is the fraction of friendships between people in the county with at least one other mutual friend in the county. Spectral homophily is the second largest eigenvalue of the row-stochasticized network adjacency matrix, a measure of the extent to which the county-level friendship network is fragmented into separate groups. The Penn State Index is an index of participation in civic organizations and other measures of civic engagement obtained from Rupasingha et al. 2006. Civic organizations is the number of civic organizations with Facebook pages per 1,000 Facebook users in the county. Volunteering rate is the percentage of Facebook users in the county who are members of “volunteering” or “activism” groups. See Supplementary Table 2 for an expanded version of this correlation table that includes all social capital measures that we construct. See Section VI.B for further details on all the social capital measures.

TABLE 3: Correlations between Social Capital Measures and Upward Income Mobility

	County Level		ZIP Code Level	
	(1)	(2)	(1)	(2)
(1) Economic Connectedness (EC)	0.65***	(0.04)	0.69***	(0.02)
(2) EC Restricted to Top 10 Friends	0.64***	(0.04)	0.69***	(0.02)
(3) Mean Friend Rank for Indivs. at p=25	0.65***	(0.04)	0.66***	(0.02)
(4) Bottom-to-Top SES Quintile EC	0.56***	(0.06)	0.55***	(0.04)
(5) Facebook Childhood EC	0.44***	(0.04)		
(6) Instagram Childhood EC	0.62***	(0.05)		
(7) Spectral SES Homophily	-0.21***	(0.08)	-0.02	(0.06)
(8) Age Connectedness	-0.16***	(0.05)	-0.24***	(0.03)
(9) Language Connectedness	0.06	(0.06)	0.09***	(0.03)
(10) Clustering	0.03	(0.05)	0.14***	(0.04)
(11) Support Ratio	-0.23***	(0.06)	-0.05	(0.03)
(12) Spectral Homophily	-0.11	(0.06)		
(13) Penn State Index	0.06	(0.06)		
(14) Local Trust Index	0.11	(0.08)		
(15) Volunteering Rate	0.18***	(0.05)	0.26***	(0.02)
(16) Civic Organizations	0.07	(0.05)	-0.02*	(0.01)

Notes: This table reports county-level and ZIP-code-level univariate correlations between upward income mobility and the full set of social capital measures we construct in this paper, expanding upon the subset shown in Figure 4a. All correlations are weighted by the number of children born to parents with below-median income as reported in the Opportunity Atlas (Chetty et al. 2018). Standard errors (reported in parentheses) are clustered by commuting zone for county-level correlations and by county for ZIP code-level correlations. For definitions of the variables used in this table, see the notes to Figure 4; Sections II.A, II.B, II.C, and VI; and Supplementary Information A.5.3 and B.2. Asterisks indicate the level of significance: *5%, **1%, ***0.1%.

TABLE 4: Associations between Race-Specific Upward Income Mobility and Economic Connectedness in Racially Homogeneous Areas

Upward Mobility for:	White Individuals				Black Individuals		Hispanic Individuals	
	Counties		ZIP Codes		ZIP Codes		ZIP Codes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economic Connectedness	0.68*** (0.04)	0.69*** (0.05)	0.69*** (0.02)	0.69*** (0.02)	0.61*** (0.13)	0.65*** (0.16)	0.36* (0.15)	0.31 (0.21)
Sample	>80% White	>90% White	>80% White	>90% White	>80% Black	>90% Black	>80% Hispanic	>90% Hispanic
Observations	1,955	1,427	16,087	12,058	247	105	174	63
Focal Race Share in Sample	90%	95%	91%	95%	90%	95%	89%	94%

Notes: This table presents correlations between race-specific measures of upward income mobility and economic connectedness across racially homogeneous counties and ZIP codes. Data on upward income mobility by race are obtained from Chetty et al. 2020. Upward mobility is measured as the predicted household income rank in adulthood for children in the 1978-83 birth cohorts with parents at the 25th percentile of the national income distribution. In columns 1-4, we correlate upward mobility for white individuals with economic connectedness in counties and ZIP codes where over 80% or over 90% of the population is white (based on data from the 2000 Census). In columns 5-6, we correlate upward mobility for Black individuals with economic connectedness in ZIP codes where over 80% or over 90% of the population is Black. In columns 7-8, we correlate upward mobility for Hispanic individuals with economic connectedness in ZIP codes where over 80% or over 90% of the population is Hispanic. Because the statistics on upward mobility are constructed using individuals who grew up in the U.S. and a large share of Hispanic individuals are immigrants, in columns 7-8 we measure economic connectedness using only individuals who have a U.S. hometown in the Facebook data. The bottom row of the table shows the percentage of individuals in the estimation sample of the focal racial group (e.g., percentage white in columns 1-4). For white and Black individuals, the correlations are weighted by number of children with below-median parental income as calculated in the Opportunity Atlas (Chetty et al. 2018) using Census data. In columns 7-8, the weights are the number of children who have a U.S. hometown and have below-median parental SES in the Facebook data. Standard errors (reported in parentheses) below each correlation are clustered at the commuting zone level for county-level correlations and at the county level for ZIP code-level correlations. Asterisks indicate the level of significance: *5%, **1%, ***0.1%.

TABLE 5: Associations between Upward Mobility, Economic Connectedness, and Other Neighborhood Characteristics

A. Economic Connectedness vs. Median Income and Poverty Rates

Dependent Variable:	Upward Mobility							
	Counties				ZIP Codes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Median Income	0.345*** (0.045)	-0.006 (0.069)			0.574*** (0.023)	0.209*** (0.029)		
Poverty Rate			-0.299*** (0.070)	0.142** (0.069)			-0.543*** (0.052)	-0.195*** (0.054)
Economic Connectedness		0.649*** (0.058)		0.732*** (0.043)		0.548*** (0.038)		0.568*** (0.033)
Observations	2,984	2,984	2,984	2,984	24,165	24,165	24,165	24,165
R-squared	0.119	0.418	0.089	0.430	0.330	0.496	0.295	0.496

B. Economic Connectedness vs. Segregation and Inequality

Dependent Variable:	Upward Mobility					
	(1)	(2)	(3)	(4)	(5)	(6)
Income Segregation	-0.173*** (0.053)	-0.071 (0.054)				
Racial Segregation			-0.212** (0.088)	-0.027 (0.086)		
Income Inequality (Gini coefficient)					-0.449*** (0.084)	-0.103 (0.091)
Economic Connectedness		0.601*** (0.044)		0.604*** (0.054)		0.577*** (0.063)
Observations	1,820	1,820	1,821	1,821	2,741	2,741
R-squared	0.034	0.413	0.051	0.408	0.207	0.424

C. Economic Connectedness vs. Share of Black Residents

Dependent Variable:	Upward Mobility for Black Individuals				Upward Mobility for White Individuals			
	Counties		ZIP Codes		Counties		ZIP Codes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share Black	-0.158** (0.068)	0.078 (0.076)	-0.204*** (0.057)	-0.014 (0.071)	-0.128** (0.057)	0.151** (0.067)	-0.250*** (0.018)	0.035* (0.018)
Economic Connectedness		0.502*** (0.095)		0.468*** (0.083)		0.582*** (0.051)		0.631*** (0.027)
Observations	1,885	1,885	11,147	11,147	2,982	2,982	24,020	24,020
R-squared	0.025	0.222	0.042	0.224	0.016	0.277	0.063	0.380

Notes: This table presents estimates from OLS regressions of upward income mobility on economic connectedness and other area-level characteristics. Upward income mobility is obtained from the Opportunity Atlas (Chetty et al. 2018) and is measured as the predicted household income rank in adulthood for children in the 1978-83 birth cohorts with parents at the 25th percentile of the national income distribution. Economic connectedness is twice the share of above-median-SES friends among below-median-SES people. We standardize every dependent and independent variable to have a mean of zero and variance of one (weighted by the number of children with below-median parental income in the county). In Panels A and B, the dependent variables are upward mobility pooling all racial and ethnic groups (Chetty et al. 2018), and regressions are weighted by the number of children with below-median parental income. Panel A presents regressions at both the county and ZIP code levels, with median household income and poverty rates by county and ZIP code obtained from the 2000 Census. In Panel B, all regressions are estimated at the county level. Income segregation is defined using a Theil (entropy) index from Reardon and Bischoff 2011, while racial segregation is defined using Theil's H-index across four groups (white, Black, Hispanic, other); see Supplementary Information A.5.1 for details. Gini coefficients are defined as the raw Gini coefficient estimated using tax data minus the income share of the top 1% to obtain a measure of inequality among the bottom 99% in each county (Chetty et al. 2014). Panel C presents regressions at both the county and ZIP code levels. The dependent variables are upward mobility estimates for Black and white individuals separately (Chetty and Hendren 2018b). Share Black is from the 2000 Census. All regressions in this panel are weighted by the race-specific number of children with below-median parental income in the county. See Supplementary Information A.5 for further details on data sources for neighborhood-level characteristics. Standard errors (reported in parentheses) are clustered at the commuting zone level. Asterisks indicate the level of significance: *10%, **5%, ***1%.

TABLE 6: 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)	34.0	7.8	26	29	33	38	42	34.2	5.8
Male	46.4%	49.9%						50.2%	50.0%
English Language	92.9%	25.8%						94.4%	23.0%
Years on Facebook	11.3	4.1	4.3	9.9	12.5	13.9	15.4		
Number of Facebook Friends	568.2	601.0	93	182	382	737	1,251		
College Graduation Rate (County-Level)	57.9%	15.5%	38.1%	47.7%	56.2%	67.3%	79.0%	58.1%	15.7%
Median Household Income (\$, County-Level)	57,787	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 = 72.2 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	538	903	1,399		
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, and have a non-missing residential ZIP code. Panel B replicates Panel A for the sample used to measure childhood economic connectedness—users whom we can link to parents with valid SES predictions (see Supplementary Information A.2.) and assign a high school (see Supplementary Information A.3). 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. 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.

SUPPLEMENTARY TABLE 1: Typology of Social Capital Measures

		Data Used to Construct Measures		
		Labeled Network	Unlabeled Network	Non-Network
Type of Social Capital	Cross-Type Connectedness (Bridging Capital)	Economic Connectedness Age Connectedness Language Connectedness Spectral SES Homophily		Income Segregation Racial Segregation
	Network Cohesiveness		Clustering Support Ratio Spectral Homophily	
	Civic Engagement			Volunteering Rate Civic Organizations Penn State Index Local Trust Index

Notes: This table summarizes the various social capital measures we construct in the paper. The measures are divided into three types of social capital, shown in the three rows: cross-type connectedness (or bridging capital), network cohesiveness, and civic engagement. We then classify the measures by the type of data used to construct them in the three columns: network data with labels, unlabeled network data, and non-network data. The measures that we focus on in this paper are shown in black; the racial and income segregation measures, which use non-network data to measure connectedness between different types of groups, shown in grey, are analyzed in Table 5. This table covers only the measures we analyze in this paper and does not provide a comprehensive classification of many other measures of social capital discussed in prior work.

SUPPLEMENTARY TABLE 2: Correlations of Social Capital Measures Across Counties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Economic Connectedness	(1) Economic Connectedness (EC)	1.00														
	(2) EC Restricted to Top 10 Friends	0.99	1.00													
	(3) Mean Friend Rank for Indivs. at p=25	0.98	0.97	1.00												
Connectedness on Other Dimensions	(4) Top/Bottom Quintile EC	0.74	0.75	0.81	1.00											
	(5) Childhood EC	0.61	0.62	0.57	0.39	1.00										
	(6) Instagram Childhood EC	0.82	0.82	0.80	0.60	0.56	1.00									
	(7) Spectral SES Homophily	-0.24	-0.23	-0.21	-0.21	-0.02	-0.22	1.00								
Network Cohesiveness	(8) Age Connectedness	-0.45	-0.44	-0.46	-0.40	-0.35	-0.31	-0.49	1.00							
	(9) Language Connectedness	0.10	0.08	0.07	-0.08	-0.12	0.11	-0.40	0.17	1.00						
Civic Engagement	(10) Clustering	0.01	0.01	-0.01	-0.05	-0.14	0.04	-0.55	0.51	0.38	1.00					
	(11) Support Ratio	-0.25	-0.27	-0.29	-0.39	-0.18	-0.23	-0.33	0.50	0.30	0.64	1.00				
	(12) Spectral Homophily	-0.09	-0.07	-0.05	0.08	0.12	-0.14	0.59	-0.49	-0.37	-0.61	-0.51	1.00			
Civic Engagement	(13) Penn State Index	0.31	0.29	0.30	0.12	0.11	0.22	-0.07	-0.04	0.08	0.39	0.28	-0.25	1.00		
	(14) Local Trust Index	0.35	0.33	0.28	0.01	0.09	0.31	-0.24	-0.04	0.25	0.34	0.32	-0.35	0.43	1.00	
	(15) Volunteering Rate	0.46	0.44	0.44	0.16	0.16	0.39	-0.21	-0.04	0.28	0.30	0.23	-0.35	0.44	0.43	1.00
	(16) Civic Organizations	0.27	0.24	0.28	0.10	0.03	0.19	-0.15	0.05	0.16	0.37	0.23	-0.33	0.67	0.28	0.46

Notes: This table replicates Table 2 with the full set of social capital measures that we construct. See Economic Connectedness, Cohesiveness, and Civic Engagement sections of Main Text, Supplementary Information A.5.3, and Supplementary Information B.2 for definitions of each of the measures in this table.

SUPPLEMENTARY TABLE 3: Heterogeneity in ZIP Code-Level Relationships between Social Capital and Upward Income Mobility across Counties

Variables	Mean Correlation	Signal SD of Correlations	Noise SD of Correlations	% Corrs Opp. Sign of Mean
	(1)	(2)	(3)	(4)
Economic Connectedness	0.68	0.19	0.12	0.0%
Clustering	0.39	0.34	0.15	12.5%
Support Ratio	-0.08	0.26	0.16	37.6%
Volunteering Rate	0.40	0.24	0.16	4.5%

Notes: This table presents statistics summarizing the variation across counties in the correlation between upward income mobility and various social capital measures (shown in the rows) across ZIP codes within the 250 most populous counties. For each county, we first estimate the correlation between upward income mobility and the social capital measure across ZIP codes within each county, weighted by the number of children born to parents with below-median income as calculated in the Opportunity Atlas (Chetty et al. 2018) using Census data. Column 1 reports the mean value of these correlations across counties. We then estimate the heteroskedasticity-robust standard error associated with the correlation in each county. We estimate the noise variance in the correlation coefficients across counties as the weighted mean of the squared standard errors of the correlation coefficients across counties. We calculate the signal component of the variance as the total (raw) variance of the correlation coefficients across counties minus the noise variance. The Signal SD of Correlations reported in column 2 is the square root of the signal variance and is an estimate of the latent variation in the underlying (signal) distribution of the correlations in the absence of noise due to sampling error. In column 3, Noise SD of Correlations denotes the square root of the noise variance in the correlations. Column 4 reports the estimated proportion of correlations with the opposite sign to the weighted mean of these correlations, using the signal standard deviation and assuming a Normal distribution.

SUPPLEMENTARY TABLE 4: Correlations Between Economic Connectedness and Racial Shares across Counties

	(1)	(2)	(3)	(4)	(5)
(1) Economic Connectedness (EC)	1.00				
(2) Share White	0.45	1.00			
(3) Share Black	-0.43	-0.47	1.00		
(4) Share Hispanic	-0.29	-0.74	-0.16	1.00	
(5) Share Asian	0.26	-0.44	-0.04	0.28	1.00

Notes: This table reports pairwise correlations between economic connectedness and racial shares across counties, weighted by number of children with below-median parental income as calculated in the Opportunity Atlas (Chetty et al. 2018) using Census data. Economic connectedness is twice the share of above-median-SES friends among below-median-SES people in our primary analysis sample. Racial shares are obtained from the 2000 Census.

SUPPLEMENTARY TABLE 5: Variables Used to Predict Socioeconomic Status in Machine Learning Model

Number	Feature	Description
1	Age	Current year minus year of birth
2	Age at registration	Age when user created Facebook account
3	City prediction	User's home city
4	College	College attended
5	Instagram	Indicator for whether the user has an Instagram account
6	County prediction	User's predicted county
7	Days since confirmed	Number of days since Facebook account creation
8	Gender	User gender
9	Graduate school	Self-reported indicator for attendance of graduate school
10	Language	User-specified language setting
11	Phone model	Model of primary phone used
12	Estimated phone price	Current price of primary phone used
13	Region (state) prediction	User's home state
14	ZIP prediction	User's home ZIP code
15	Relationship status	Indicators for current relationship status (e.g., married)
16	Usage L28	Number of days with activity on any Facebook platform in last 28 days
17	Mobile usage L28	Number of days with activity on Facebook mobile platform in last 28 days
18	Web usage L28	Number of days with activity on Facebook web platform in last 28 days
19	Mobile carrier	Carrier of user's mobile phone
20	Phone OS type	OS of user's phone (e.g., Android, iOS)
21	Avg donation amount	Average amount that user has donated on the platform
22	Time since last donation	Time since user's last donation on the platform

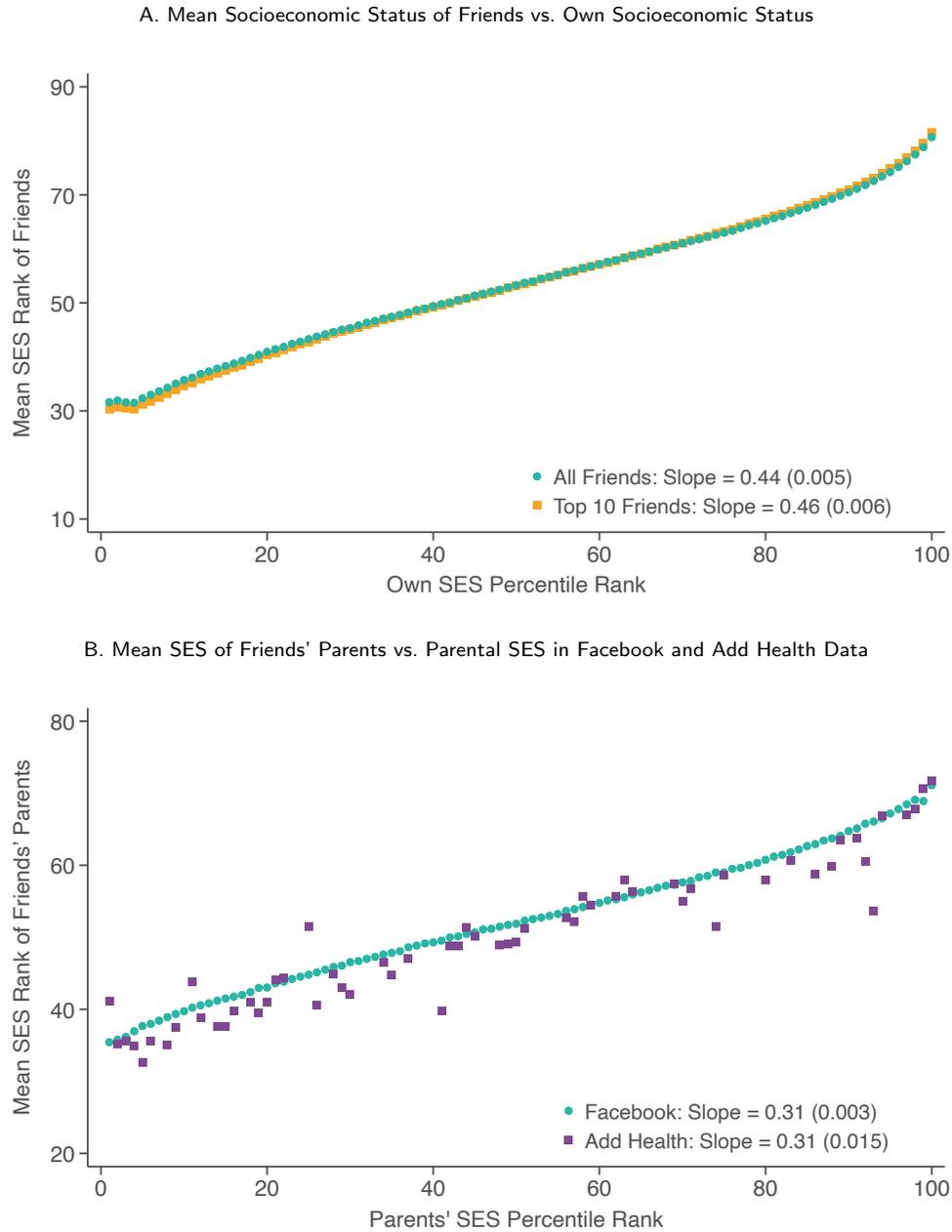
Notes: This table lists the set of variables that we use to predict individuals' socioeconomic status using the machine learning model described in Supplementary Information B.1.

SUPPLEMENTARY TABLE 6: Correlations between Economic Mobility and Measures of Economic Connectedness

	Upward Mobility (1)	Baseline ML Ests. (2)	ACS Block Group Median (3)	ACS Zip Code Median (4)	Z-Score (5)
Upward Mobility	1				
Baseline ML Estimates	0.646	1			
ACS Block Group Median	0.605	0.839	1		
ACS Zip Code Median	0.550	0.783	0.945	1	
Combined Z-Score Index	0.599	0.913	0.882	0.855	1

Notes: This table presents county-level correlations between upward income mobility and measures of economic connectedness constructed from various alternative proxies for SES described in Supplementary Information B.1. All correlations are weighted by the number of children born to parents with below-median income. Upward income mobility comes from Chetty et al. (2018), and is defined as the predicted household income rank in adulthood for children in the 1978-83 birth cohorts with parents at the 25th percentile of the national income distribution. The measures of SES we consider are: the baseline machine learning (ML) model prediction used in our main analysis; the median household income in a block group (available for the Location History subsample only) or ZIP code (available for the full primary sample) from the ACS; and a composite z-score index consisting of: (1) the number of days since account creation (an older account correlates with higher SES); (2) the price of the phone used by the individual; (3) the selectivity tier of college an individual attended; (4) the number of days out of the last 28 days that a user accessed Facebook using a web browser (more usage from a website correlates with higher SES); (5) an indicator for whether the user reports a graduate school (having a graduate school is associated with higher SES); and (6) the median household income in the user's ZIP code from the ACS. The composite index is constructed by standardizing each of these six variables (subtracting the mean, dividing by the standard deviation) and then taking an unweighted average of the six variables, with each variable signed so that higher values correlate with higher SES ML predictions.

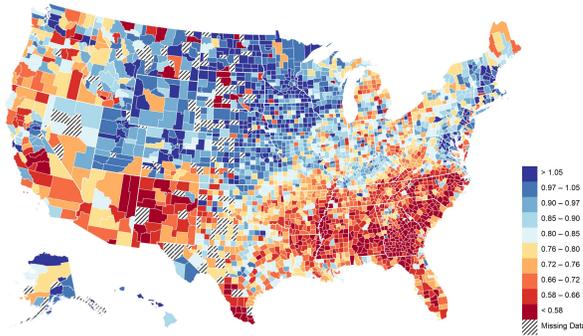
FIGURE 1: Relationship Between an Individual's SES and Friends' SES



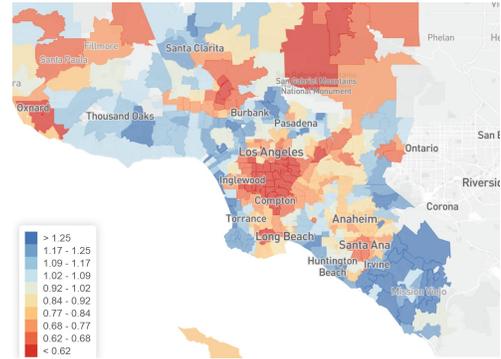
Notes: Panel A plots the mean socioeconomic status (SES) rank of individuals' friends vs. their own SES percentile ranks. The series in green circles is calculated using the entire friendship network for each individual. The series in orange squares is constructed using each individual's ten closest friends, based on the frequency of public interactions such as likes, tags, wall posts, and comments. Socioeconomic status is constructed by combining information on 22 variables to predict median household incomes in individuals' residential block groups and then ranking individuals relative to others in the same birth cohort; see Section VI.B for details. Panel B compares estimates of homophily in the Facebook data and the National Longitudinal Study of Adolescent to Adult Health (Add Health) survey. The series in purple squares plots the mean parental income rank of children's friends vs. their own parents' income percentile rank in the Add Health data. The series in green circles presents the analogous relationship in the Facebook data using our SES proxies, restricting the sample to individuals born in 1989-1994 and using their five closest high school friends to match the Add Health sample as closely as possible (see Supplementary Information A.5.2). For each series, we report slopes estimated from a linear regression on the plotted points, with heteroskedasticity-robust standard errors in parentheses.

FIGURE 2: The Geography of Social Capital in the United States

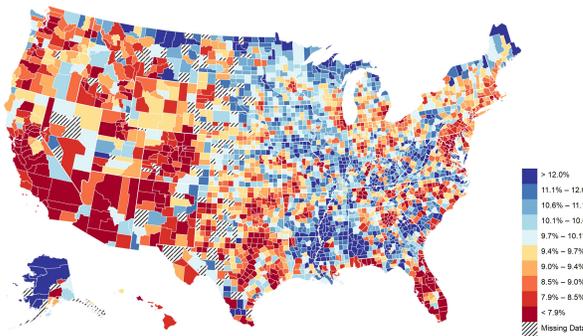
A. Economic Connectedness, by County



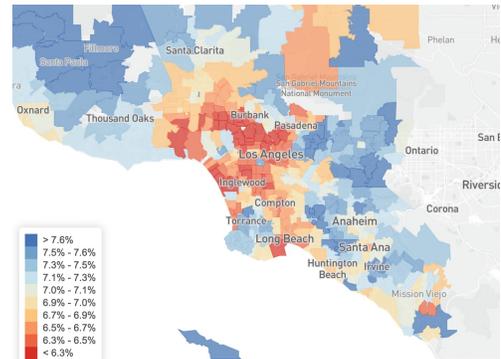
B. Economic Connectedness, by ZIP Code in Los Angeles



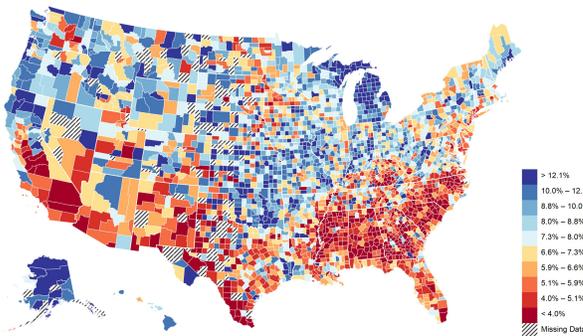
C. Clustering, by County



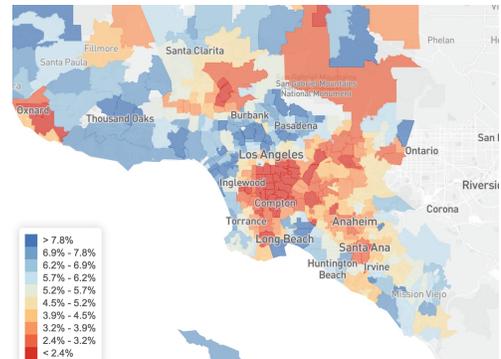
D. Clustering, by ZIP Code in Los Angeles



E. Volunteering Rate, by County

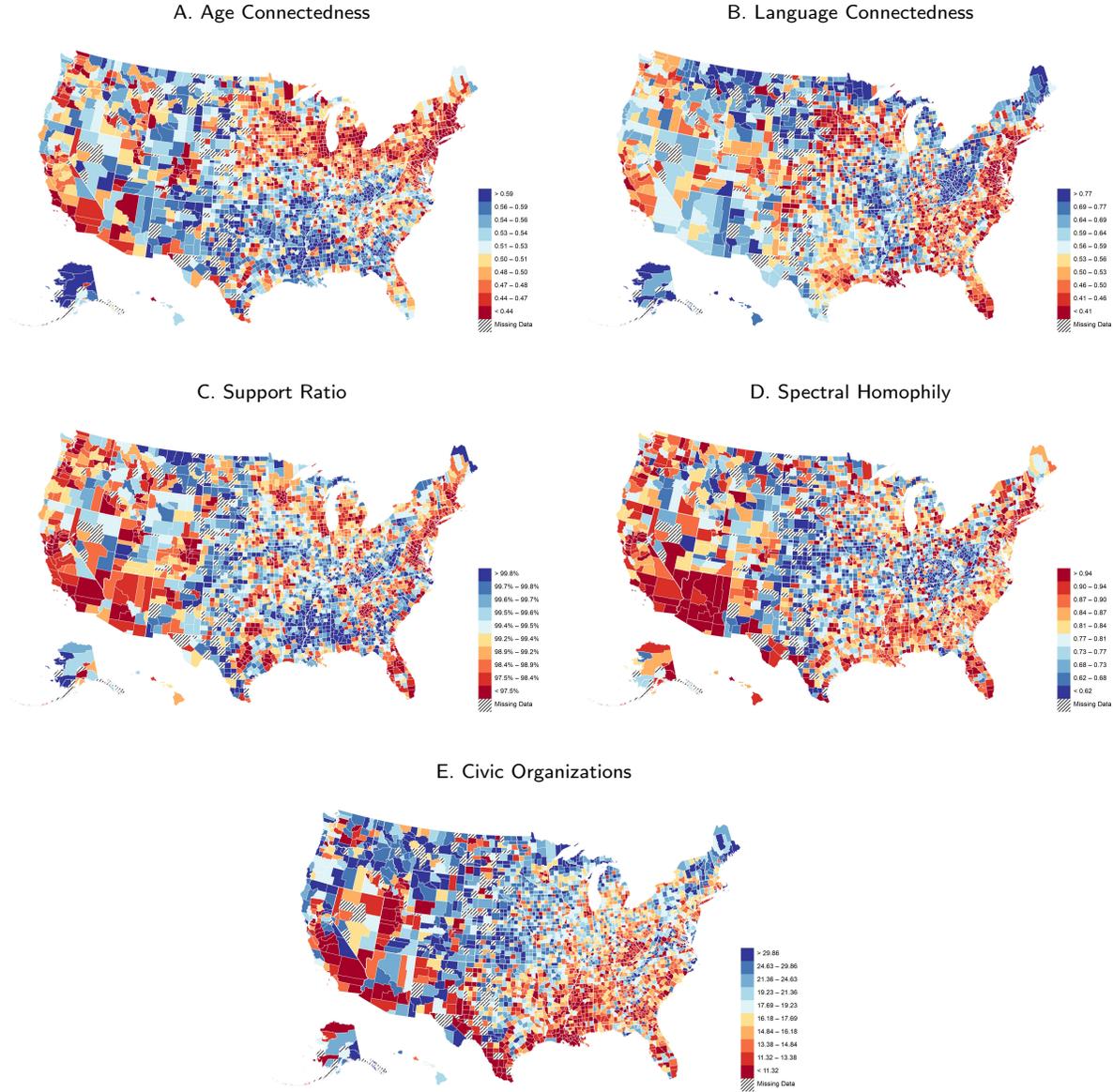


F. Volunteering Rate, by ZIP Code in Los Angeles



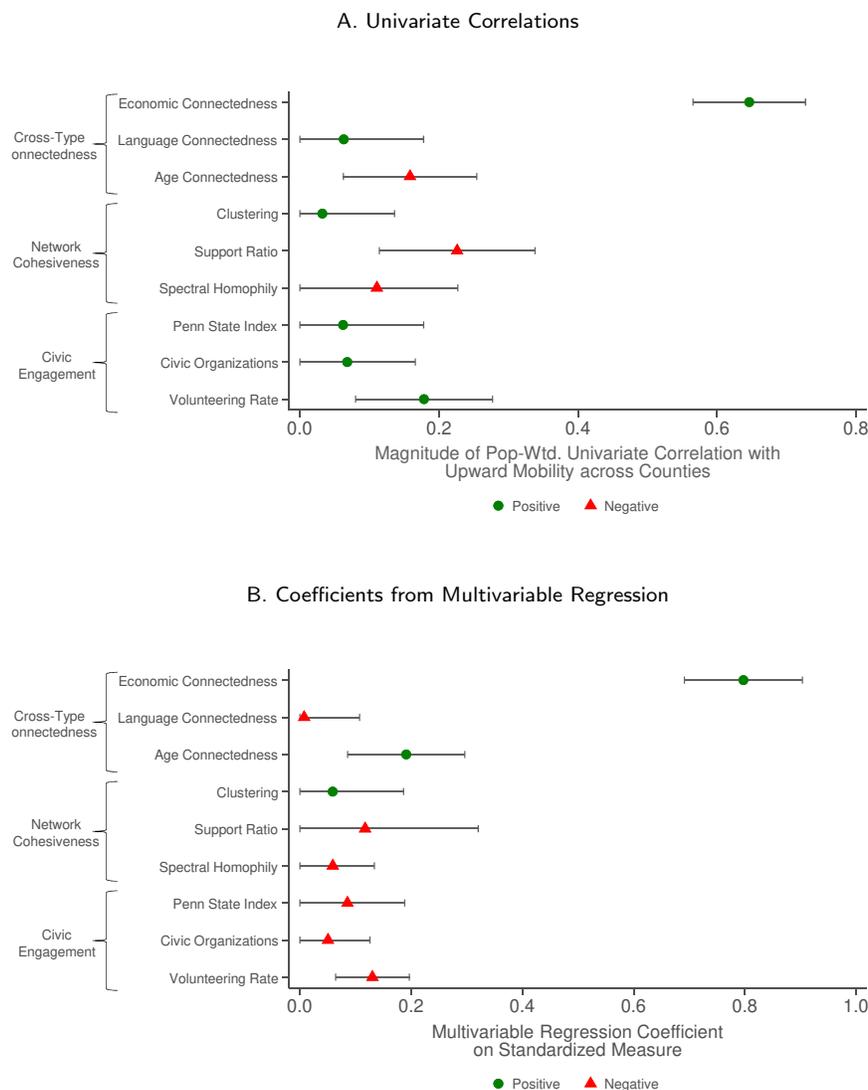
Notes: Panel A shows a county-level map of economic connectedness (EC), defined as twice the fraction of above-median-SES friends among below-median-SES people. Panel B shows a ZIP code-level map of EC in Los Angeles. Panel C shows a county-level map of average clustering, the fraction of an individual's friend pairs who are friends with each other. Panel D shows a ZIP code-level map of average clustering in Los Angeles. Panel E shows a county-level map of volunteering rates, the percentage of individuals who are members of "volunteering" or "activism" groups as classified by Facebook. Panel F shows a ZIP code-level map of volunteering rates in Los Angeles. We omit counties and ZIP codes where statistics are estimated on fewer than 50 below-median-SES Facebook users. These maps must be viewed in color to be interpretable. Analogous maps for all ZIP codes in the U.S. are available at www.socialcapital.org. Figure 3 presents county-level maps of other social capital measures.

FIGURE 3: Geographical Variation in Other Social Capital Measures



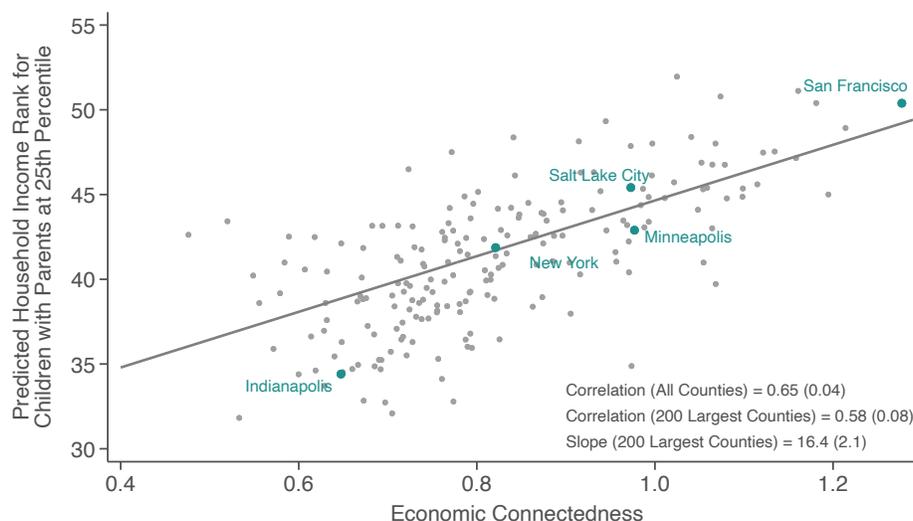
Notes: This figure presents county-level maps analogous to those in Figure 2 for other measures of social capital. These maps must be viewed in color to be interpretable. Age connectedness (Panel A) is the average share of friends who are 35 to 44 among users who are 25 to 34, normalized by the sample share of individuals who are 35 to 44. Language connectedness (Panel B) is the average share of friends who set their Facebook language to English among individuals who do not set their Facebook language to English, normalized by the sample share of people who set their Facebook language to English. Support ratio (Panel C) is the share of friendships between people in the county who have at least one other friend in the county in common. Spectral homophily (Panel D) is the second largest eigenvalue of the row-stochasticised network adjacency matrix, a measure of the extent to which the county-level friendship network is fragmented into separate groups. Civic organizations (Panel E) is the number of civic organizations with Facebook pages per 1,000 Facebook users in the county. See Sections II.A, II.B, II.C, and VI for details on the definitions and construction of these measures.

FIGURE 4: County-Level Correlations between Upward Income Mobility and Measures of Social Capital



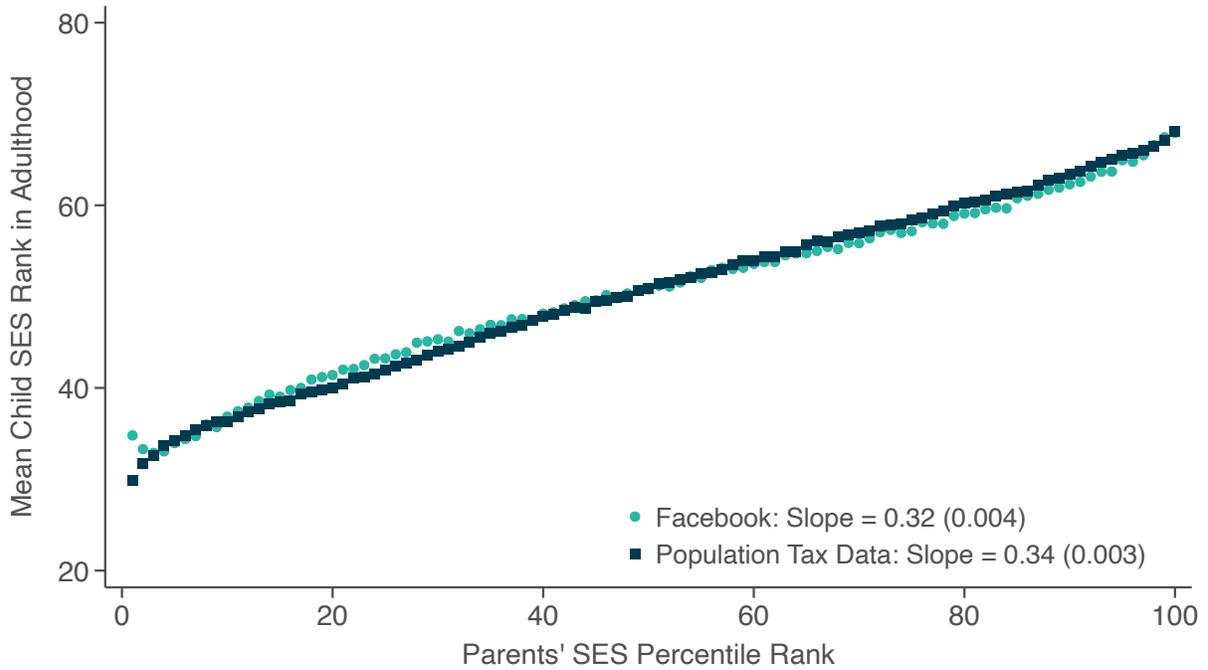
Notes: Panel A plots county-level univariate correlations of upward income mobility with social capital measures. Table 3 lists the correlation coefficients plotted here. Panel B presents estimates from a multivariable regression of upward mobility on all variables in Panel A together, standardizing the outcome and dependent variables to have mean 0 and standard deviation 1. Upward income mobility is obtained from the Opportunity Atlas (Chetty et al. 2018) and is measured as the predicted household income rank in adulthood for children in the 1978-83 birth cohorts with parents at the 25th percentile of the national income distribution. Economic connectedness is twice the share of above-median-SES friends among below-median-SES people. Language connectedness is the share of friends who set their Facebook language to English among users who do not set their language to English, divided by the national share of users who set their language to English. Age connectedness is the share of friends who are 35 to 44 among users who are 25 to 34, divided by the national share of users aged 35-44. Clustering is the fraction of an individual’s friend pairs who are also friends with each other, averaged over all individuals in the county. Support ratio is the fraction of friendships between people in the county with at least one other mutual friend in the county. Spectral homophily is the second largest eigenvalue of the row-stochasticized network adjacency matrix, a measure of the extent to which the county-level friendship network is fragmented into separate groups. The Penn State Index is an index of participation in civic organizations and other measures of civic engagement obtained from Rupasingha et al. 2006. Civic organizations is the number of civic organizations with Facebook pages per 1,000 Facebook users in the county. Volunteering rate is the percentage of Facebook users in the county who are members of “volunteering” or “activism” groups. All correlations and regressions are weighted by the number of children in each county whose parents have below-national-median income. Intervals represent 95% confidence intervals calculated using standard errors clustered by commuting zone.

FIGURE 5: Association between Upward Income Mobility and Economic Connectedness across Counties



Notes: This figure presents a scatter plot of upward income mobility vs. economic connectedness for the 200 most populous U.S. counties. Economic connectedness is defined as twice the fraction of above-median-SES friends among below-median-SES individuals living in the county. Upward income mobility is obtained from the Opportunity Atlas (Chetty et al. 2018) and is measured as the predicted household income rank in adulthood for children in the 1978–83 birth cohorts with parents at the 25th percentile of the national income distribution. We report a slope estimated using an ordinary least squares (OLS) regression on the 200 largest U.S. counties by population, with standard errors clustered by commuting zone in parentheses. We also report the population-weighted correlation between upward mobility and economic connectedness across both the 200 largest counties as well as all counties, with standard errors (clustered by commuting zone) in parentheses. The correlations and regression are weighted by the number of children in each county whose parents have below-national-median income.

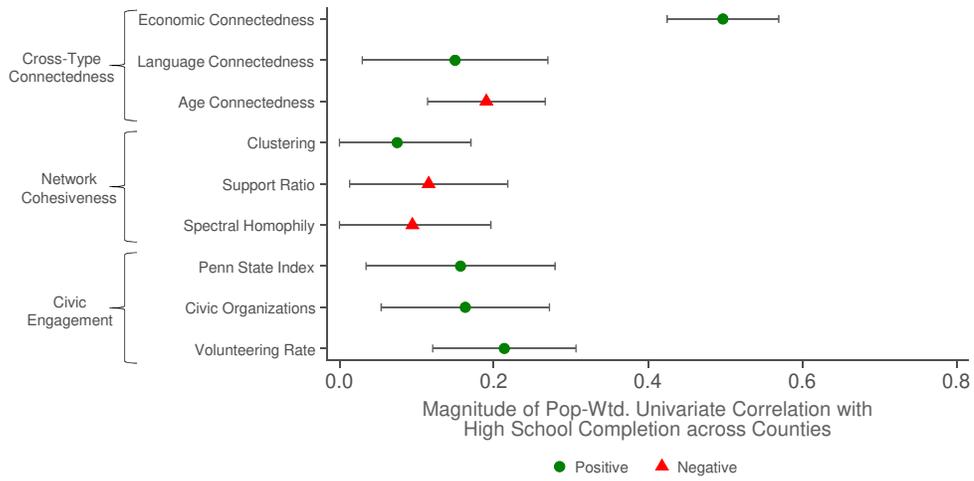
FIGURE 6: Intergenerational Persistence of Socioeconomic Status in Facebook and Tax Data



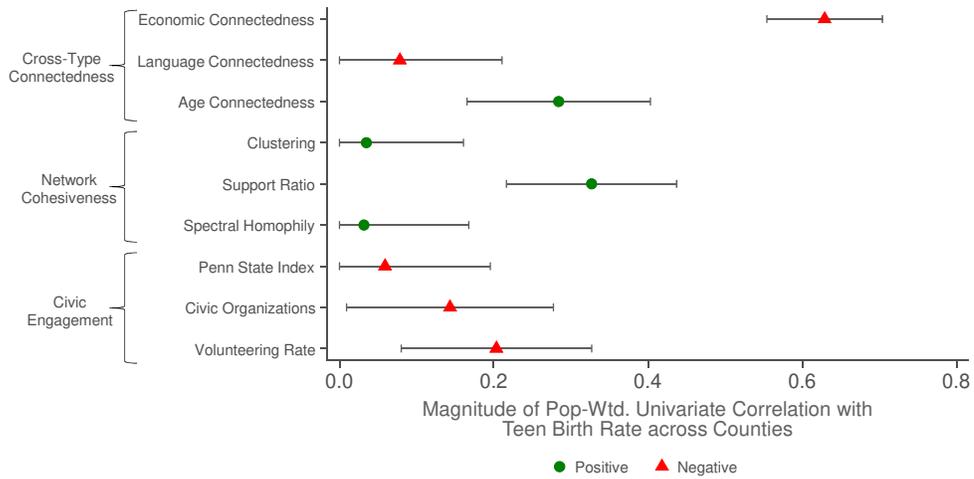
Notes: This figure shows binned scatter plots of children’s mean SES ranks in adulthood vs. their own parents’ SES ranks. Each point plots the mean SES rank of children who have parents at a given percentile of the SES distribution. The series in circles is based on data from Facebook, with SES rank calculated as described in Section VI.B. The series in squares is based on administrative tax data analyzed by Chetty et al. 2017, with SES ranks corresponding to household income ranks. The sample for both series is children born between 1980 and 1982. In both samples, children’s SES ranks are based on their ranks within their birth cohort among children linked to parents, while parents’ SES ranks are based on their ranks relative to other parents in the same group of parents linked to children born between 1980-82. We report a slope estimated using a linear regression for each series, with heteroskedasticity-robust standard errors in parentheses.

FIGURE 7: County-Level Univariate Correlations between Other Outcomes and Measures of Social Capital

A. High School Completion Rate for Children with Parents at 25th Percentile

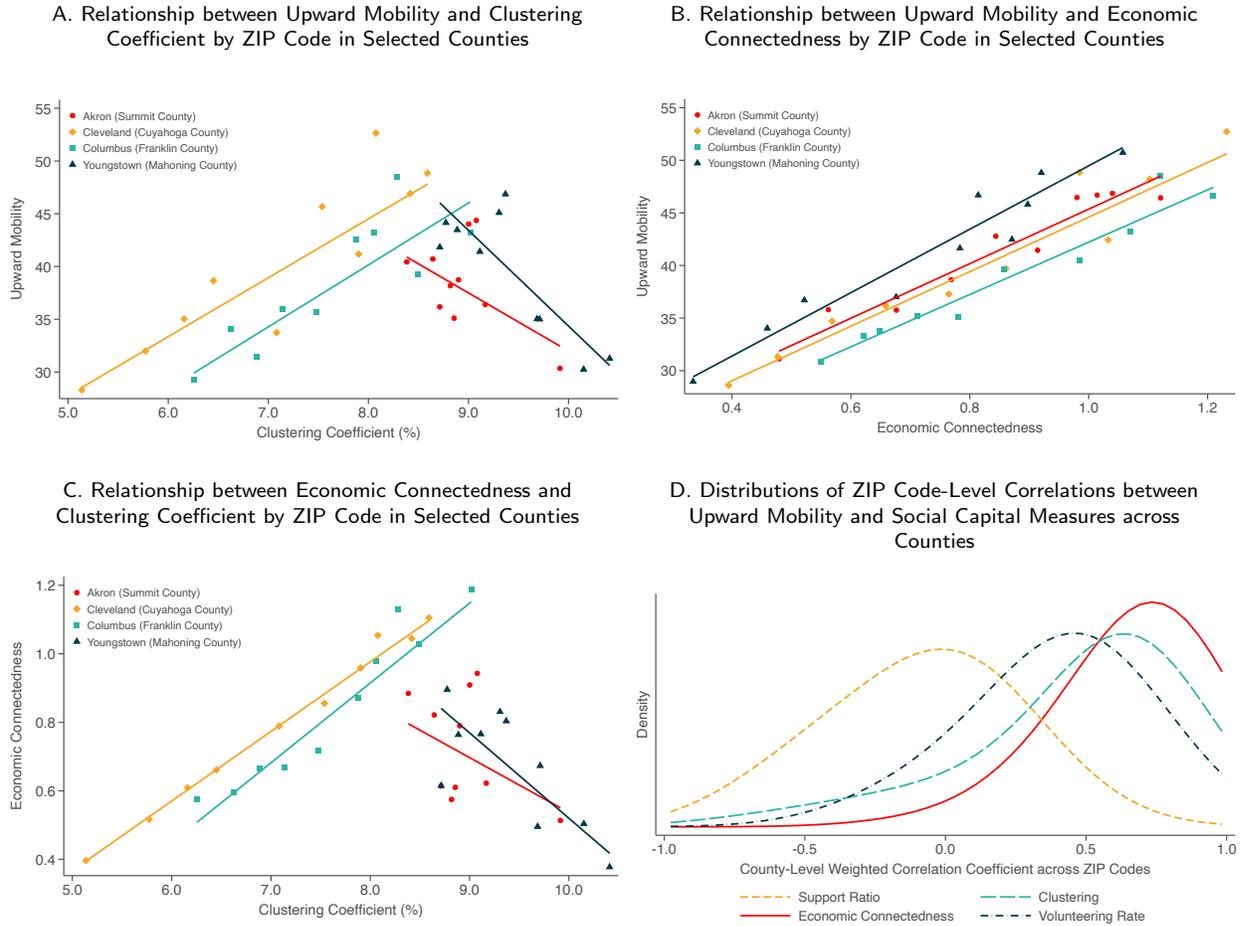


B. Teen Birth Rate for Women with Parents at 25th Percentile



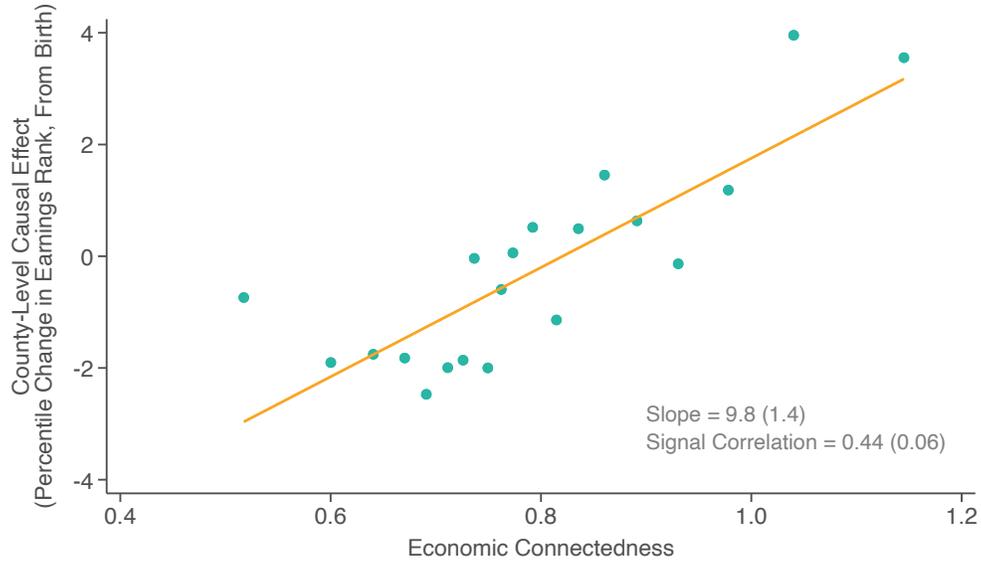
Notes: This figure replicates the across-county correlations shown in Figure 4a with two different outcome variables: high school completion rates (Panel A) and teen birth rates (Panel B) for children with parents at the 25th percentile of the national income distribution. These outcome variables are obtained from the Opportunity Atlas (Chetty et al. 2018). See notes to Figure 4 for further details.

FIGURE 8: Heterogeneity in Relationships between Upward Income Mobility and Social Capital Measures across Counties



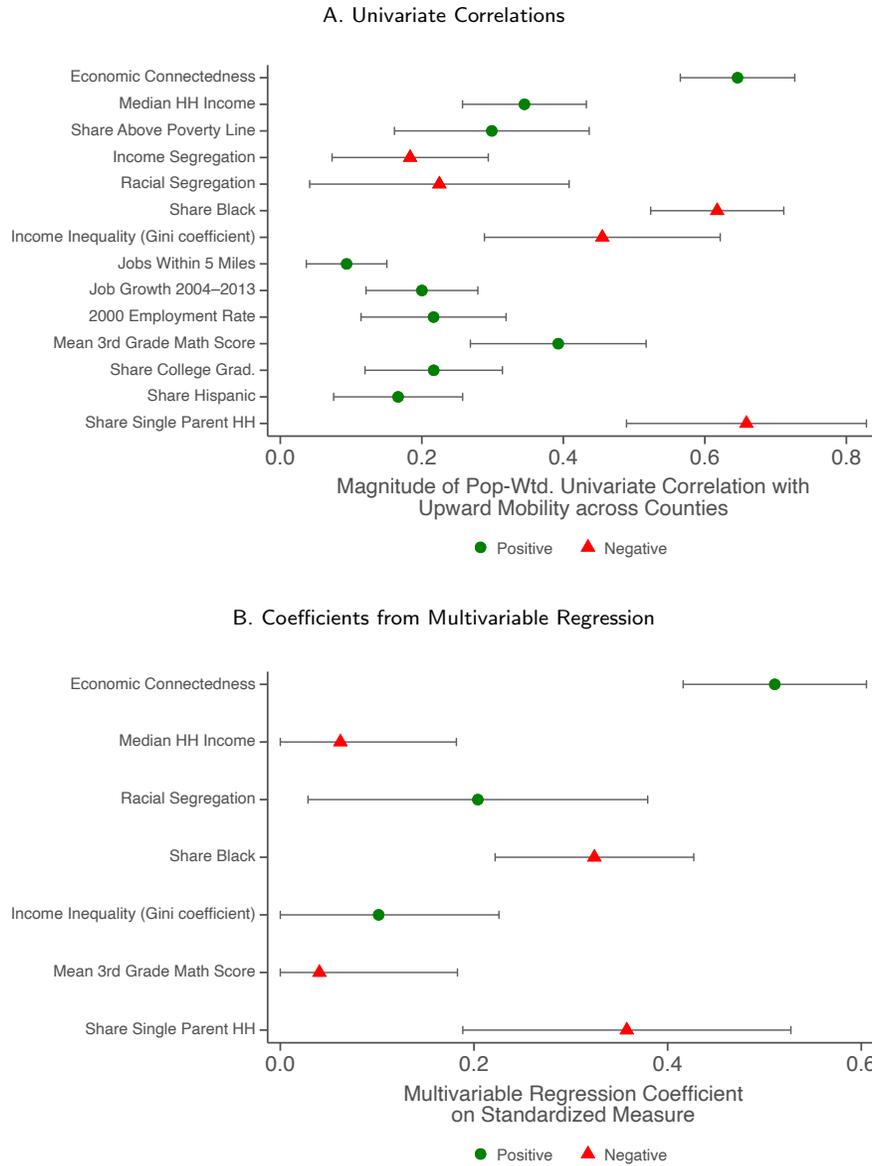
Notes: Panel A presents ZIP code-level binned scatter plots of upward income mobility vs. the degree of clustering in networks across ZIP codes in four counties in Ohio: Summit County (Akron), Cuyahoga County (Cleveland), Franklin County (Columbus), and Mahoning County (Youngstown). Clustering is defined as the share of an individual’s friend pairs who are also friends with each other, averaged over all individuals in a ZIP code. Panel B presents analogous ZIP code-level binned scatter plots of upward mobility vs. economic connectedness. Panel C presents ZIP code-level binned scatter plots of economic connectedness vs. clustering coefficients. To construct these binned scatter plots, we group ZIP codes within each county into ten (population-weighted) bins based on the relevant social capital measure shown on the x axis and plot the mean (population-weighted) level of the outcome variable vs. the social capital measure within each bin. Panel D presents kernel density plots of the distribution of ZIP-code-level correlations between upward mobility and several social capital measures across counties for the 250 most populous counties. To construct these distributions, we first estimate correlations between upward mobility and the social capital measure of interest at the ZIP code level in each county, and then plot the distribution of these correlations. All correlations and distributions are weighted by the number of children whose parents earn less than the national median household income in each ZIP code and county, respectively.

FIGURE 9: Association between Economic Connectedness and Counties' Causal Effects on Upward Income Mobility



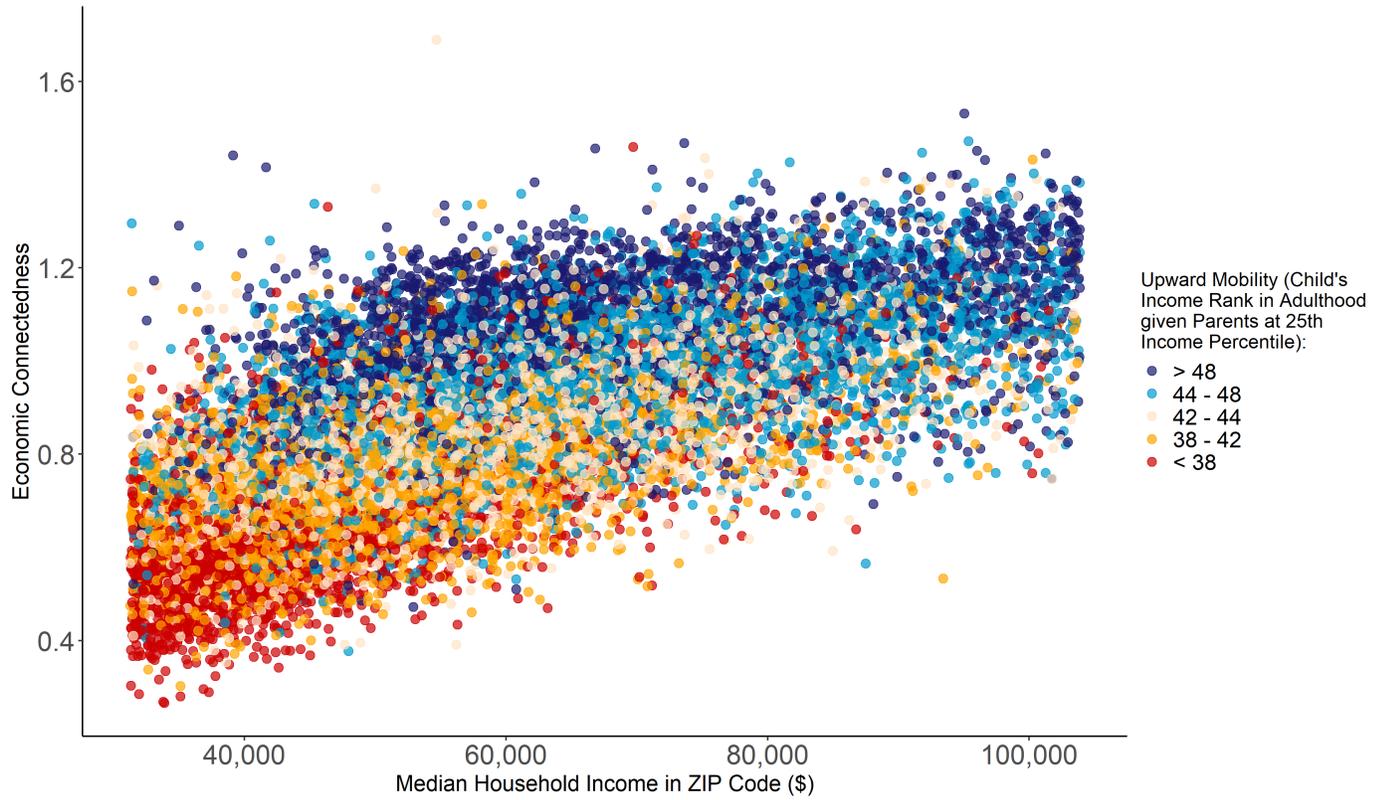
Notes: This figure presents a binned scatter plot of counties' causal effects on upward mobility vs. economic connectedness. The binned scatter plot is constructed in the same way as described in the notes to Figure 8 using 20 bins of Economic Connectedness instead of 10 and weighting by the precision (inverse of standard error squared) of the causal effect estimates. Causal effects on upward mobility are the annual exposure effect estimates constructed by Chetty and Hendren 2018b by analyzing cross-county movers. These annual exposure effects are multiplied by 20 so that they can be interpreted as the causal effect of growing up in a given location from birth to age 20 on an individual's household income percentile rank in adulthood. The slope is estimated using an OLS regression of the causal effect estimates on EC, weighting by the precision of the causal effect estimates. The signal correlation is calculated by dividing the raw (precision-weighted) correlation between the causal effects and EC by the square root of the precision-weighted reliability of the estimated causal effects.

FIGURE 10: County-Level Correlations between Upward Income Mobility and Neighborhood Characteristics



Notes: Panel A plots county-level univariate correlations of upward income mobility with economic connectedness and other county characteristics obtained from external datasets (see Supplementary Information A.5 for details). Upward income mobility is obtained from the Opportunity Atlas (Chetty et al. 2018) and is measured as the predicted household income rank in adulthood for children in the 1978–83 birth cohorts with parents at the 25th percentile of the national income distribution. Income segregation is defined using a Theil (entropy) index from Reardon and Bischoff 2011, while racial segregation is defined using Theil’s H-index across four groups (white, Black, Hispanic, other); see Supplementary Information A.5.1 for details. Gini coefficient is obtained from Chetty et al. 2014, and is defined as the raw Gini coefficient estimated using tax data minus the income share of the top 1% to obtain a measure of inequality among the bottom 99% in each county. The rest of the variables are all obtained from the Opportunity Atlas (Chetty et al. 2018). In Panel B, we present estimates from a single multivariable regression of upward mobility on a subset of variables from Panel A, with both the outcome and dependent variables standardized to have mean 0 and standard deviation 1. The variables used in Panel B are the seven variables from Panel A that have the largest univariate correlations with upward mobility (except the share of households above the poverty line, which is highly correlated with median household incomes), which include all of the strongest predictors of mobility identified by prior work (Chetty et al. 2014). All correlations and regressions are weighted by the number of children in each county whose parents have below-national-median income. Intervals represent 95% confidence intervals calculated using standard errors clustered by commuting zone.

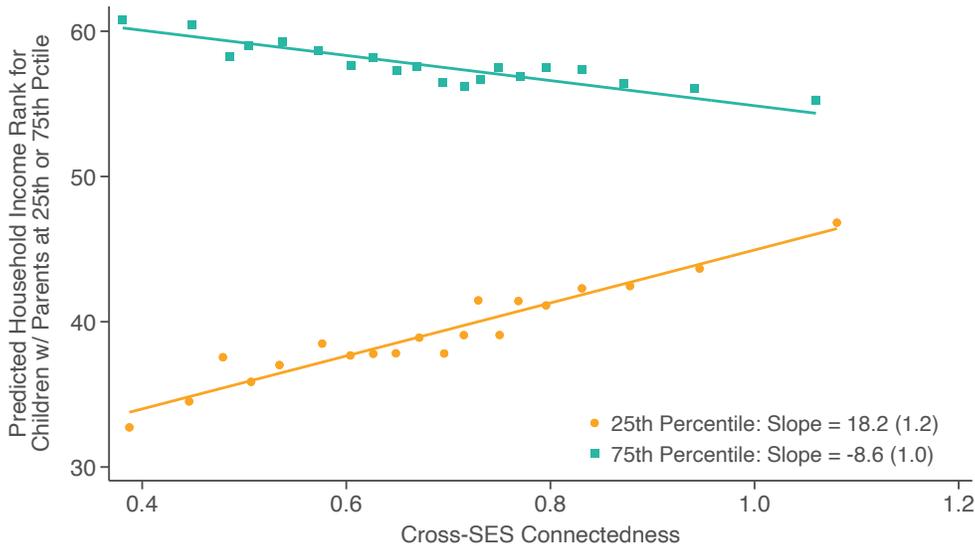
FIGURE 11: Associations Between Upward Income Mobility, Economic Connectedness, and Median Household Income, by ZIP Code



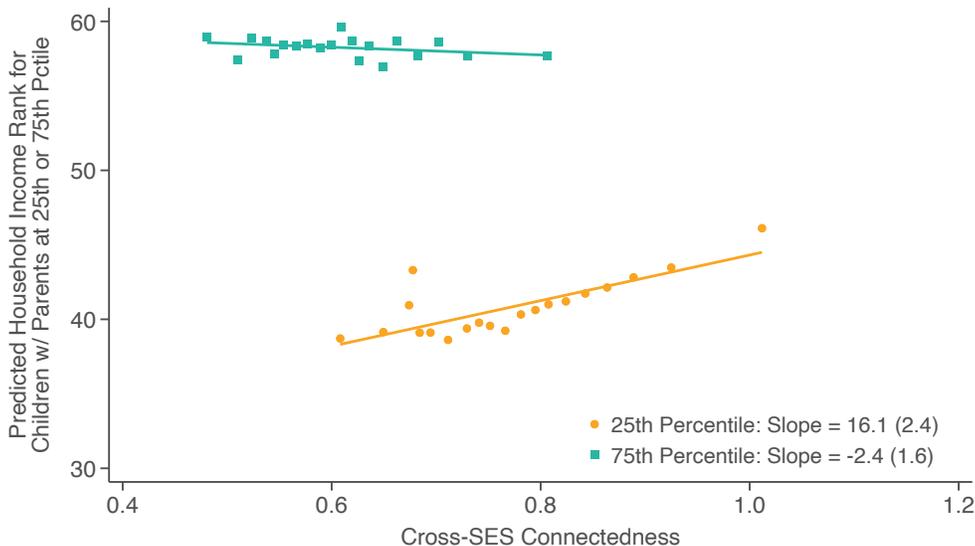
Notes: This figure presents a scatter plot of economic connectedness vs. median household income (based on the 2014–18 American Community Survey) by ZIP code. Economic connectedness is defined as twice the fraction of above-median-SES friends among below-median-SES individuals. The points are colored by the level of upward income mobility for children who grow up in the ZIP code. Upward income mobility is obtained from the Opportunity Atlas (Chetty et al. 2018) and is measured as the predicted household income rank in adulthood for children in the 1978–83 birth cohorts with parents at the 25th percentile of the national income distribution.

FIGURE 12: Associations between Upward Income Mobility and Economic Connectedness for Low- vs. High-SES Individuals

A. Upward Income Mobility vs. Cross-SES Connectedness for Low- vs. High-SES Individuals

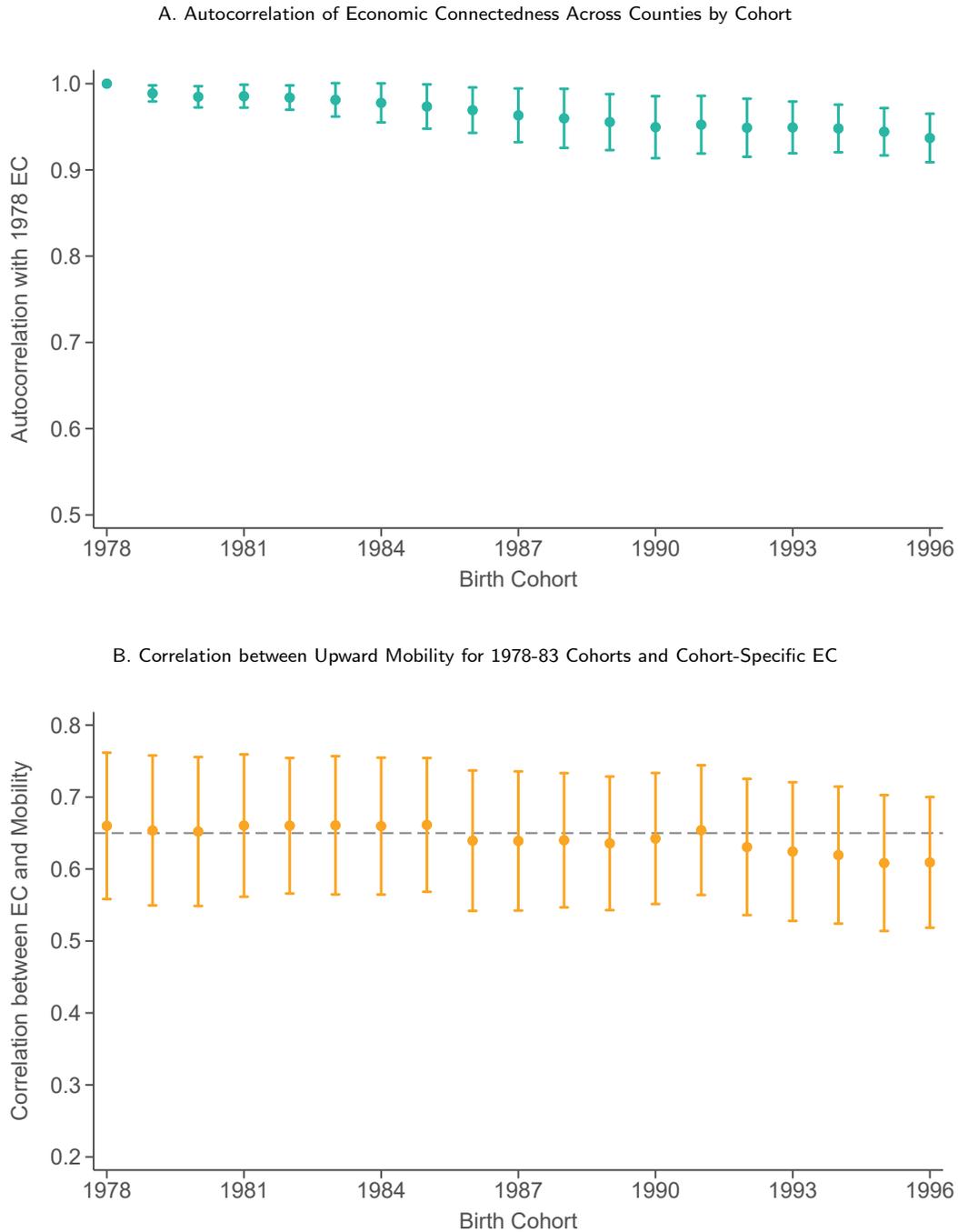


B. Upward Income Mobility vs. Cross-SES Connectedness for Low- vs. High-SES Individuals, Controlling for High-SES Shares



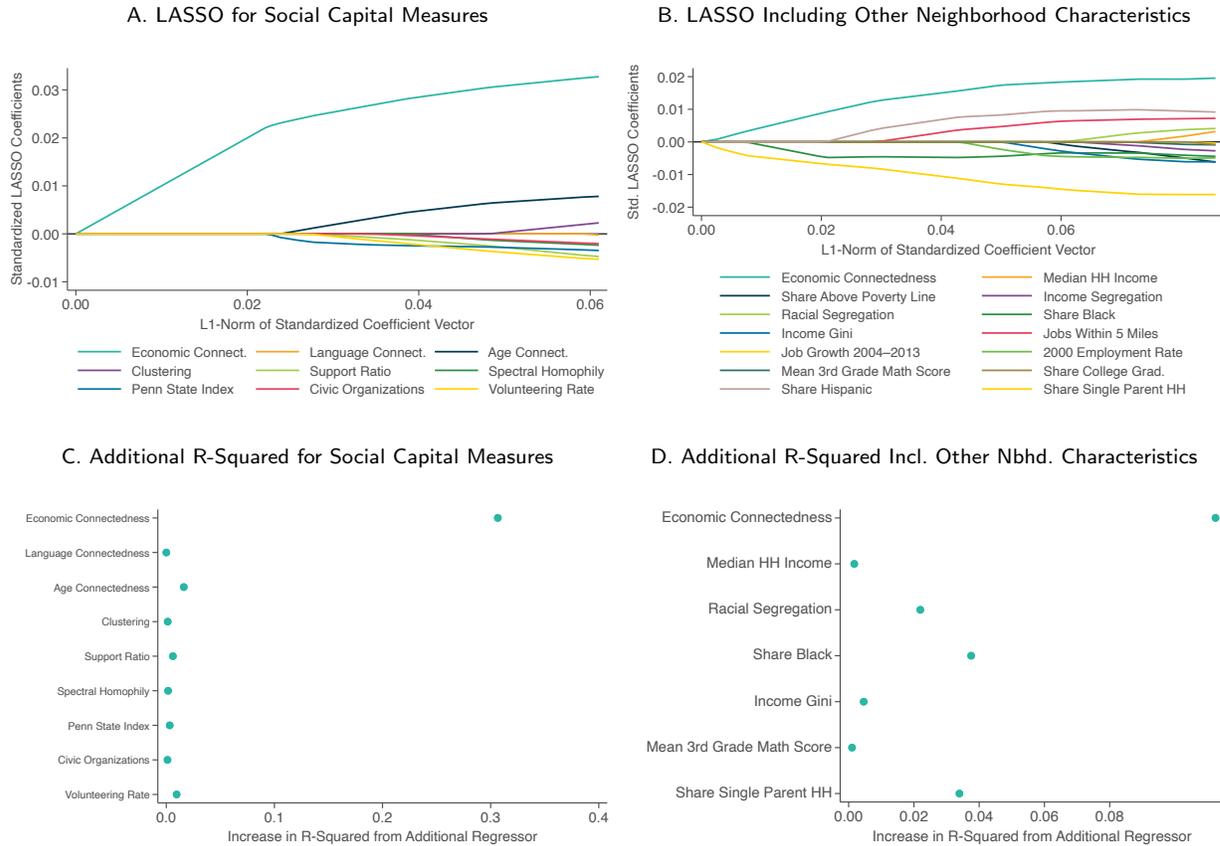
Notes: This figure presents binned scatter plots of children’s predicted income ranks in adulthood vs. cross-SES connectedness by county, separately for children with low-income (25th percentile) parents and high-income (75th percentile) parents. Data on children’s outcomes are obtained from the Opportunity Atlas (Chetty et al. 2018). We define cross-SES connectedness as the normalized share of friends for an individual in one SES group who belong to the other SES group. For below-median SES individuals, cross-SES connectedness is the same as our baseline definition of economic connectedness. Hence, the series in orange circles in Panel A is a binned scatter plot analog of Figure 5, pooling data from all counties (see notes to Figure 8 for details on construction of binned scatter plots). For above-median-SES individuals, cross-SES connectedness is twice the share of their friends who are low-SES. Panel B replicates Panel A, controlling for the share of high-SES individuals in each county. The series in Panel B are constructed by first residualizing predicted household income ranks and cross-SES connectedness on the share of high-SES people using univariate OLS regressions, and then constructing a binned scatter plot of the residuals after adding back the means of each variable for scaling purposes. We report estimates of the slope of each series based on OLS regressions with standard errors, clustered by commuting zone, in parentheses.

SUPPLEMENTARY FIGURE 1: Stability of County-Level Economic Connectedness Across Cohorts



Notes: Panel A plots the correlation across counties between economic connectedness estimated using individuals in cohort x (for x ranging from 1978 to 1996) and EC estimated using individuals in the 1978 birth cohort. Panel B plots the correlation across counties between economic connectedness estimated using individuals in cohort x and upward income mobility for the 1978–83 birth cohorts, as estimated by Chetty et al. (2018). All correlations are weighted by the number of children in each county whose parents earn less than the national median income. Vertical lines represent 95% confidence intervals estimated using standard errors clustered by commuting zone. The dashed line in Panel B shows the correlation between economic connectedness estimated on the entire primary sample and upward income mobility.

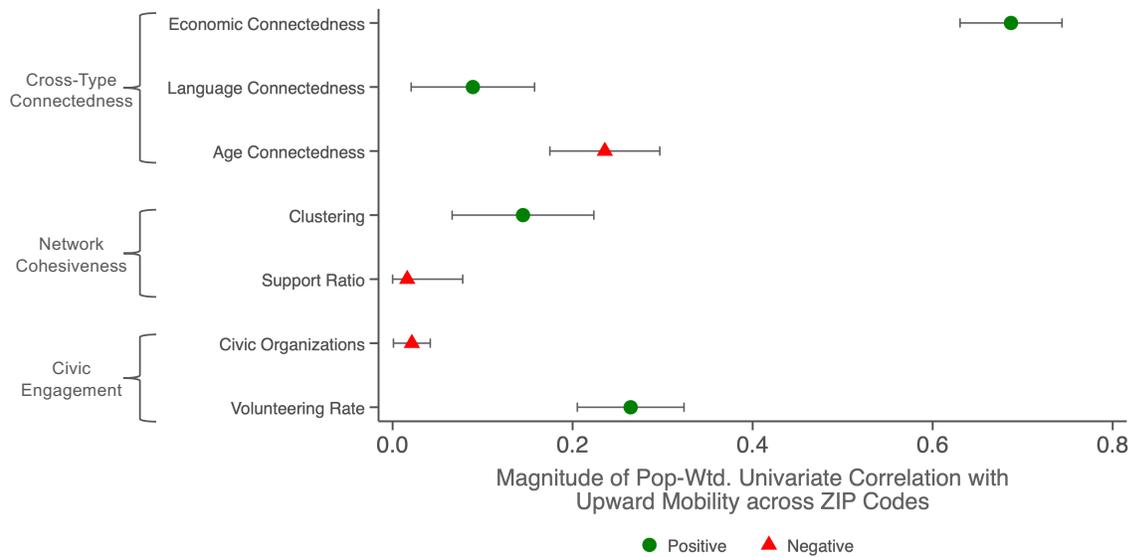
SUPPLEMENTARY FIGURE 2: LASSO Estimates & Incremental R-Squared of Predictors in Multivariable Models



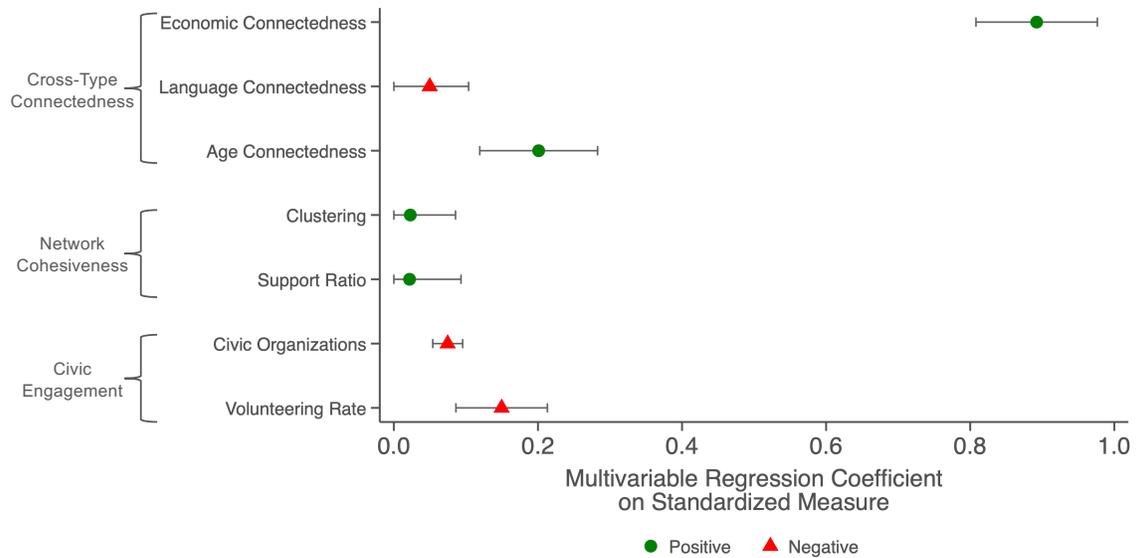
Notes: Panel A shows the standardized coefficients obtained from a LASSO regression of upward mobility on the set of social capital measures used in Figure 4a, plotted against the sum of the absolute values of the standardized coefficients. Panel B presents an analogous plot using economic connectedness and the other neighborhood characteristics used in Figure 10a. Panel C presents the incremental R-squared from adding each of the social capital measures to a regression that already includes all of the other measures used in the multivariable OLS regression specification in Figure 4b. Panel D presents an analogous plot for the regression specification in Figure 10b. All regressions are weighted by the number of children with parents who earn below the national median as reported in the Opportunity Atlas (Chetty et al. 2018) using Census data.

SUPPLEMENTARY FIGURE 3: ZIP Code-Level Correlations between Upward Mobility and Measures of Social Capital

A. Univariate Correlations



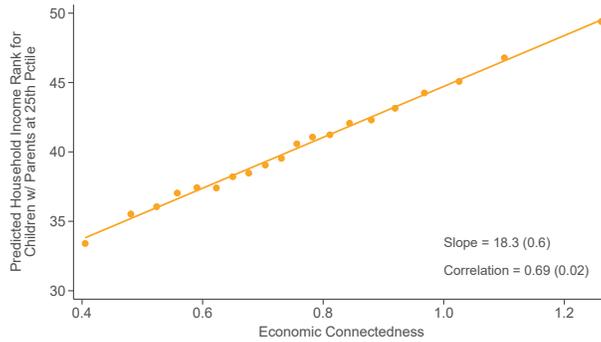
B. Coefficients from Multivariable Regression



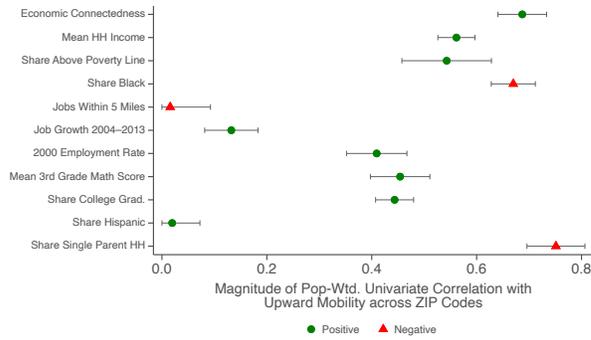
Notes: This figure replicates Figure 4 at the ZIP code level instead of the county level for the measures that are available at the ZIP code level. See note to Figure 4 for details.

SUPPLEMENTARY FIGURE 4: ZIP Code-Level Correlations between Economic Mobility and Neighborhood Characteristics

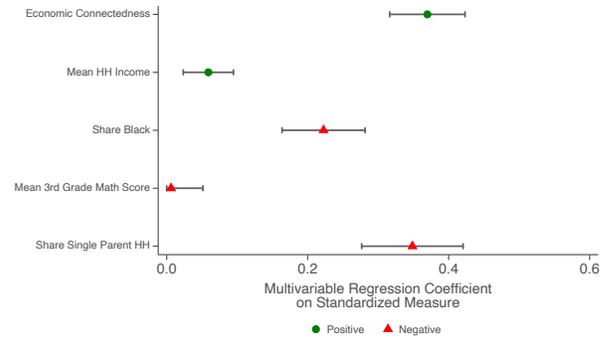
A. Relationship between Upward Mobility and EC



B. Univariate Correlations



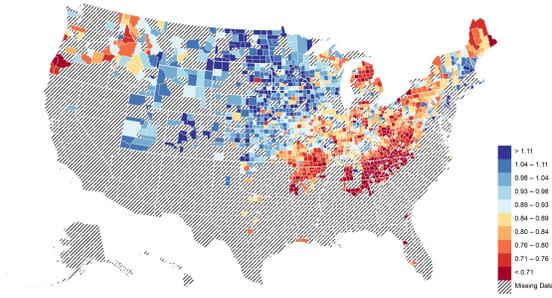
C. Coefficients from Multivariable Regression



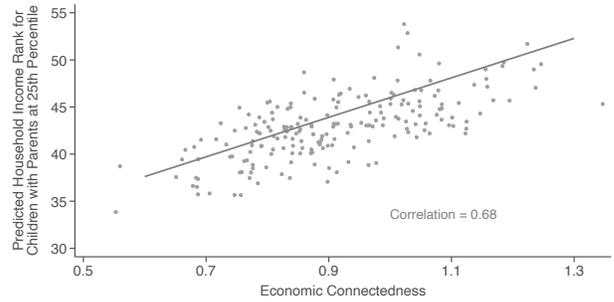
Notes: Panel A of this figure replicates the orange (below-median) series of Figure 12a at the ZIP code level. Panels B and C of this figure replicate Figure 10 at the ZIP code level instead of the county level. Income inequality (Gini coefficient) and segregation are omitted from Panels B and C because they are traditionally measured at broader geographies. The regressions are weighted by the number of children in each ZIP code whose parents have below-national-median income. Standard errors are clustered at the county level. See notes to Figure 10 and Figure 12 for further details.

SUPPLEMENTARY FIGURE 5: Social Capital and Upward Mobility in Counties with Predominantly White Residents

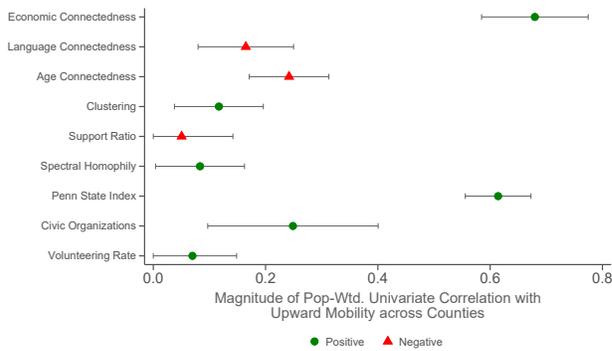
A. Spatial Variation in EC



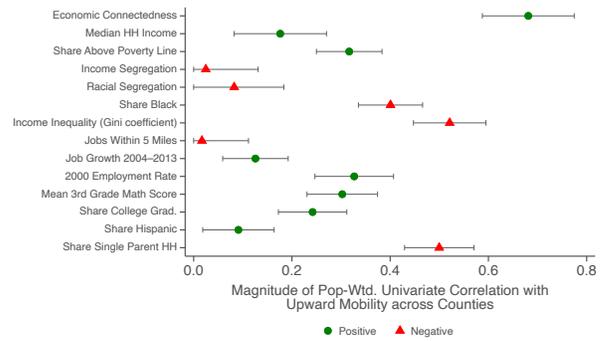
B. Relationship between Upward Mobility and EC



C. Univariate Correlations between Upward Mobility and Social Capital

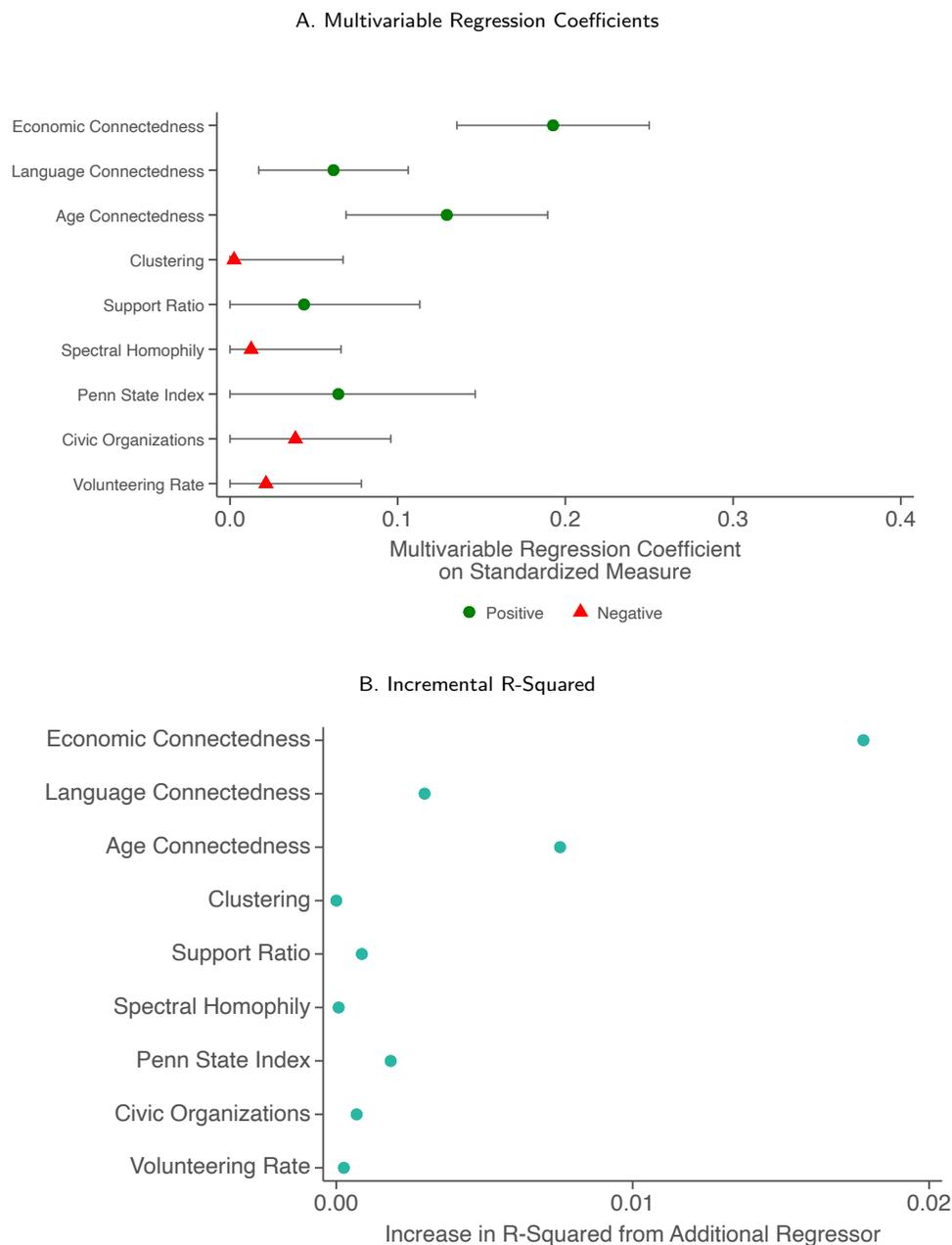


D. Univariate Correlations between Upward Mobility and Other Neighborhood Characteristics



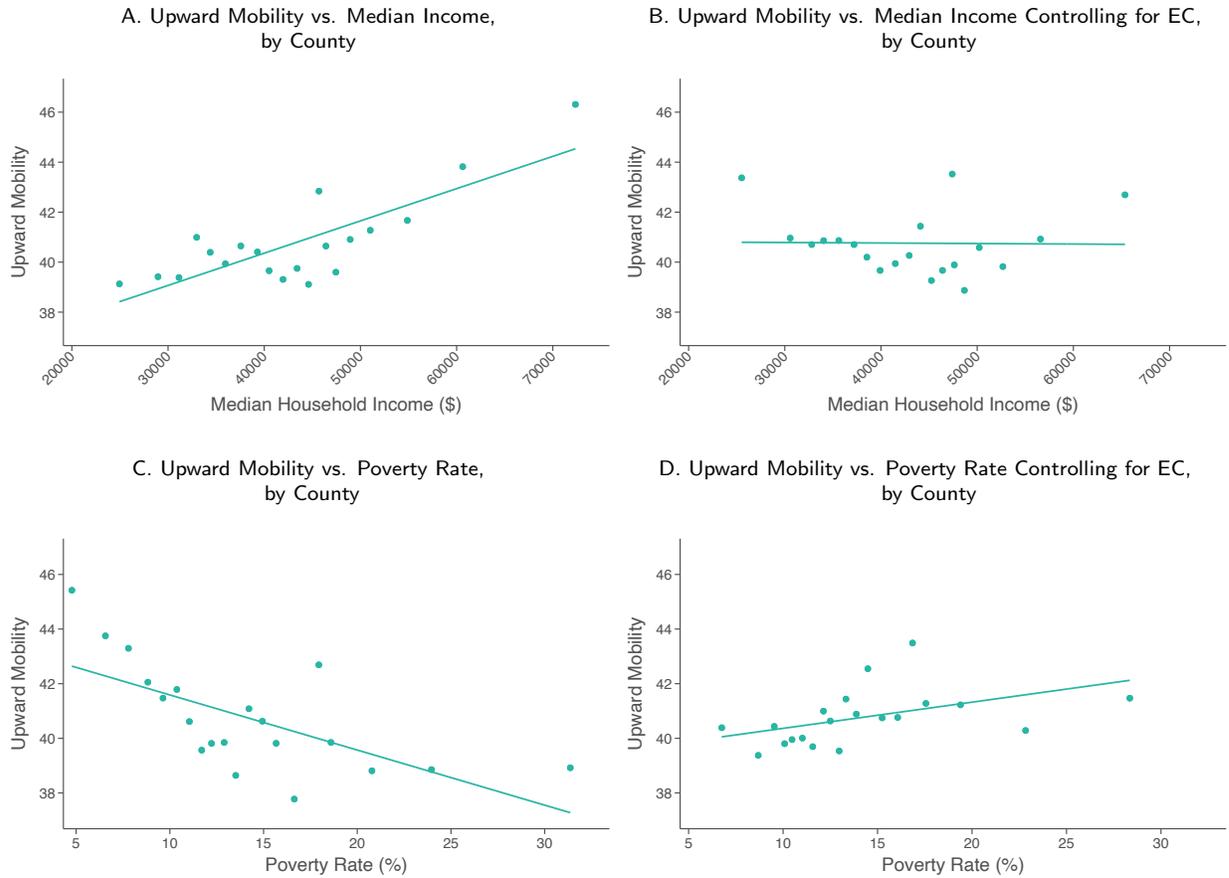
Notes: In this figure, we restrict the sample to the 1,427 counties where more than 90% of the population is white according to the 2000 Census. Panels A–D replicate Figures 2a, 5, 4a, and 10a, respectively, on this subsample of counties. See notes to those figures for further details.

SUPPLEMENTARY FIGURE 6: Regression of Counties' Causal Effects on Upward Mobility on Social Capital



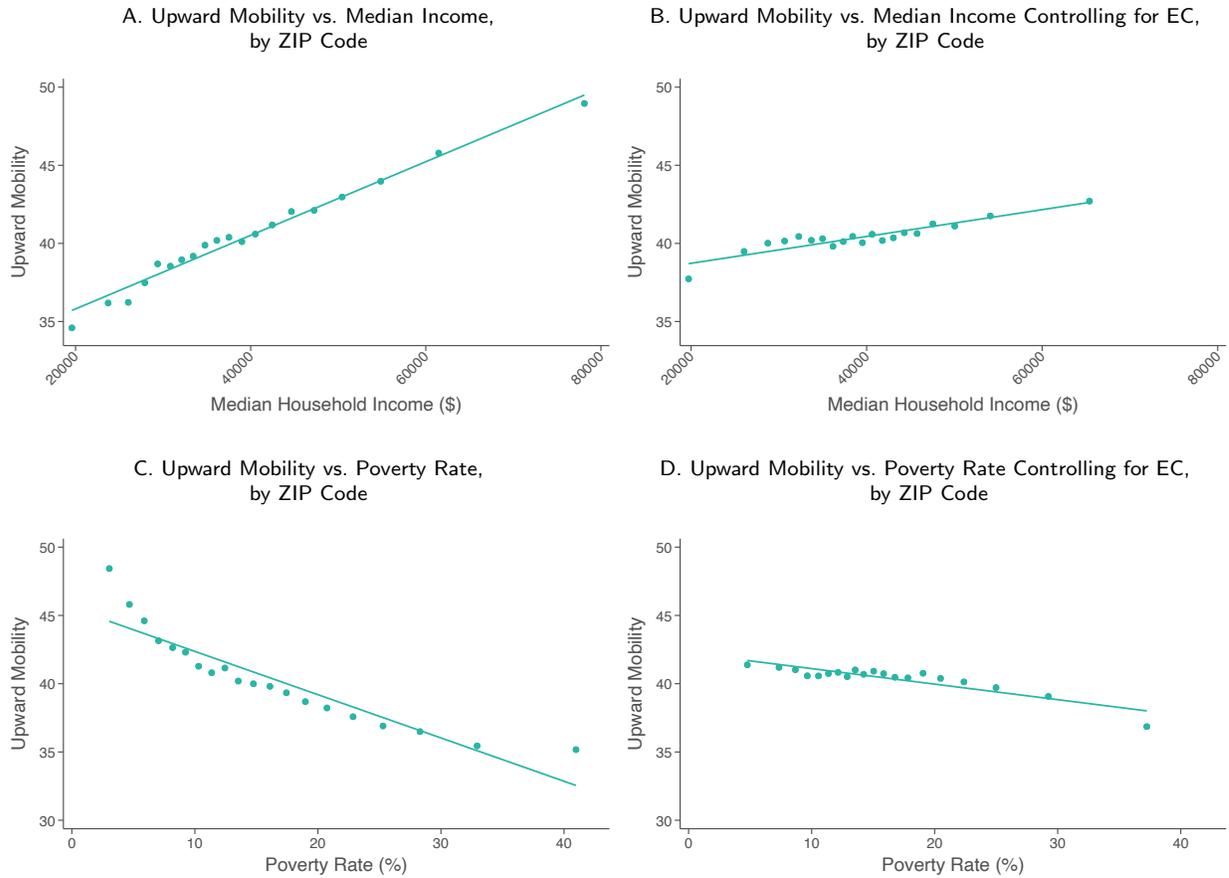
Notes: Panels A and B of this figure replicate Figure 4b and Supplementary Figure 2c, respectively, using causal instead of observational estimates of upward mobility. Causal effects on upward mobility are the raw (unshrunk) annual exposure effect estimates constructed by Chetty and Hendren (2018b) by analyzing cross-county movers. Regressions are weighted by the number of children with parents who earn below the national median as calculated in the Opportunity Atlas (Chetty et al. 2018) using Census data.

SUPPLEMENTARY FIGURE 7: Upward Mobility, Economic Connectedness, and Income Levels across Counties



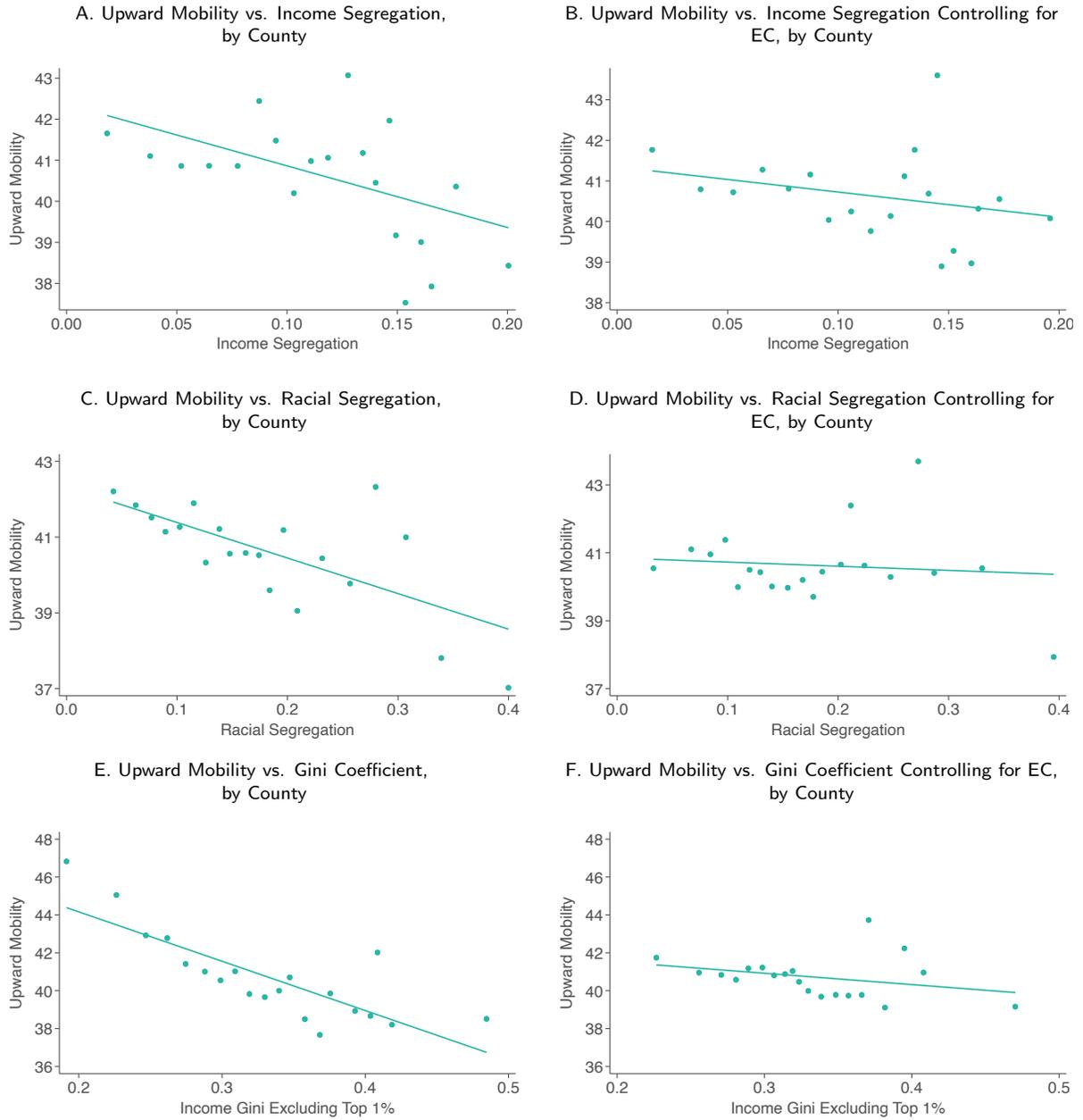
Notes: This figure shows county-level binned scatter plots of upward mobility vs. median income and poverty rates, providing non-parametric explorations of the specifications in Table 5a. The left-hand panels show raw binned scatter plots, while the right-hand panels show the same binned scatter plots controlling for economic connectedness (EC). To construct the raw binned scatter plots, we group counties into twenty bins (weighted by the number of children in each county whose parents have below-national-median income) based on values plotted on the horizontal axis (median incomes in Panels A and B, and poverty rates in Panels C and D). We then plot the (weighted) mean level of upward mobility and the income or poverty measure within each bin. The plots that control for EC are constructed by first residualizing upward mobility and the relevant horizontal-axis variable on economic connectedness using univariate OLS regressions, and then plotting a standard binned scatter plot of the residuals after adding back the means of each variable for scaling purposes. See notes to Table 5 for variable definitions and other details.

SUPPLEMENTARY FIGURE 8: Upward Mobility, Economic Connectedness, and Income Levels across ZIP Codes



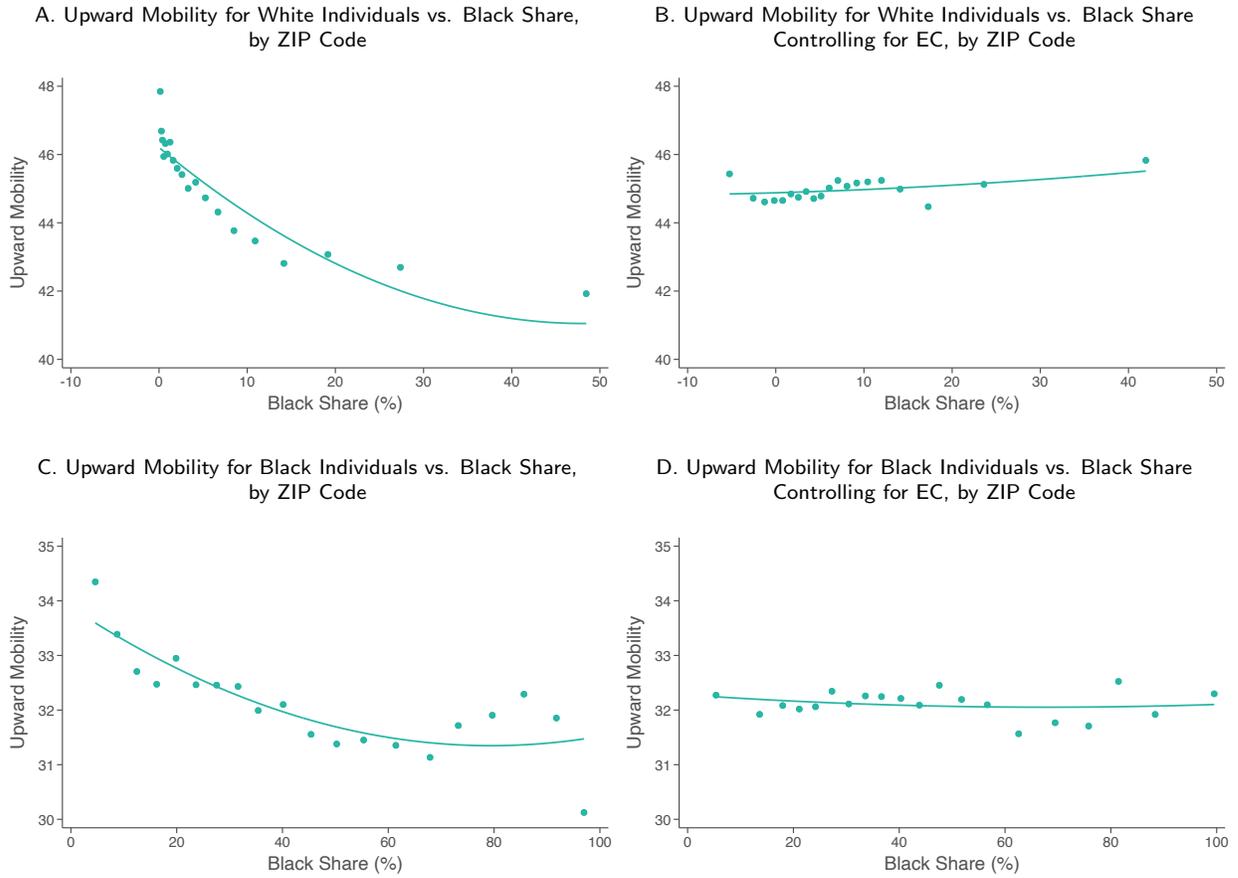
Notes: This figure presents ZIP code-level binned scatter plots of upward mobility vs. median income and poverty rates, corresponding to the linear regression specifications in Table 5a. The left-hand panels show raw binned scatter plots, while the right-hand panels show the same binned scatter plots controlling for economic connectedness (EC). See Supplementary Figure 7 for details on construction of these binned scatter plots, and the notes to Table 5 for variable definitions and other details.

SUPPLEMENTARY FIGURE 9: Upward Mobility, Economic Connectedness, and Measures of Inequality and Segregation across Counties



Notes: This figure presents county-level binned scatter plots of upward mobility vs. several measures of within-county inequality and segregation, corresponding to the linear regression specifications in Table 5b. The left-hand panels show raw binned scatter plots, while the right-hand panels show the same binned scatter plots controlling for economic connectedness (EC). See Supplementary Figure 7 for details on construction of these binned scatter plots, and the notes to Table 5 for variable definitions and other details.

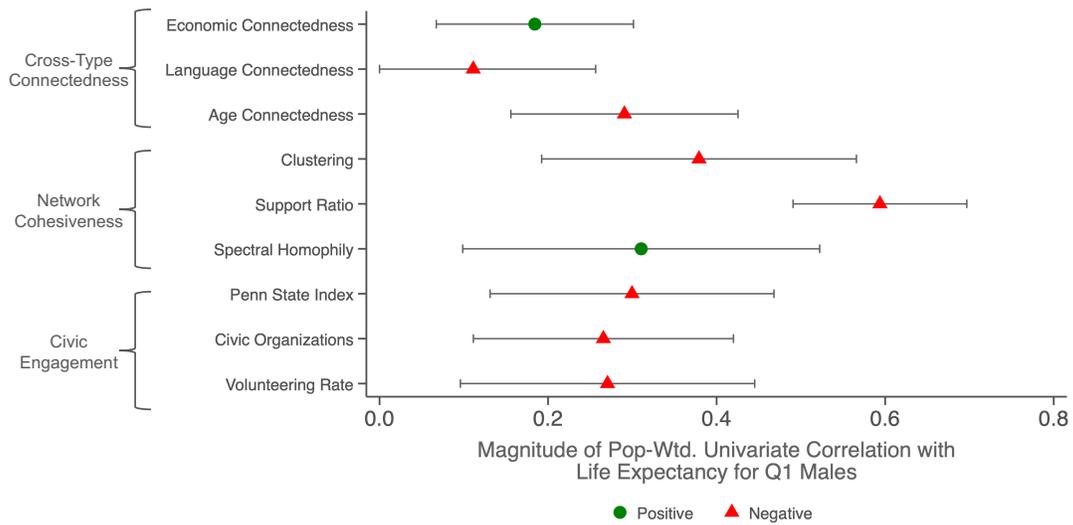
SUPPLEMENTARY FIGURE 10: Race-Specific Upward Mobility, Economic Connectedness, and Share of Black Residents across ZIP Codes



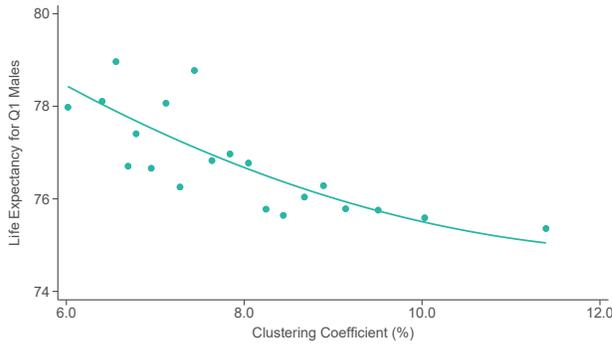
Notes: This figure presents ZIP code-level binned scatter plots of race-specific upward mobility vs. the share of Black individuals in the ZIP code, corresponding to the linear regression specifications in Table 5c. The left-hand panels show raw binned scatter plots, while the right-hand panels show the same binned scatter plots controlling for economic connectedness (EC). See Supplementary Figure 7 for details on construction of these binned scatter plots, and the notes to Table 5 for variable definitions and other details.

SUPPLEMENTARY FIGURE 11: Correlations between Social Capital and Life Expectancy

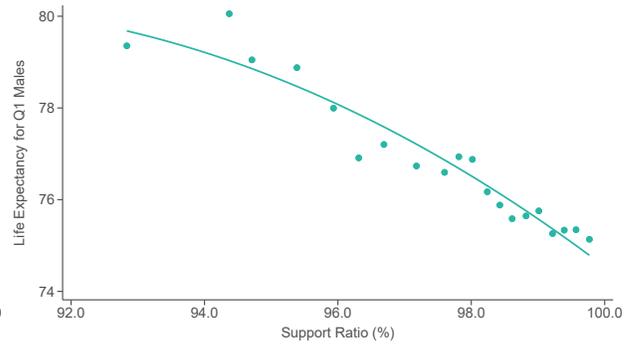
A. Correlations of Social Capital with Life Expectancy at Age 40 for Bottom-Income-Quartile Men Across Counties



B. Life Expectancy for Bottom-Income-Quartile Men vs. Clustering Coefficient, by County

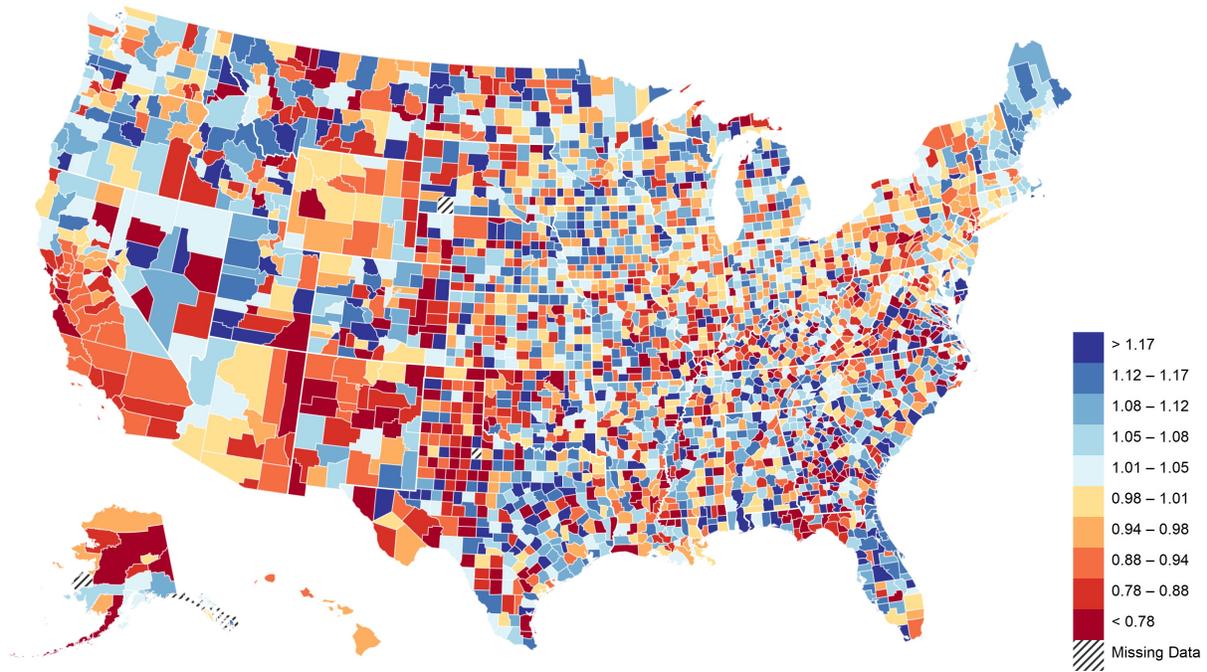


C. Life Expectancy for Bottom-Income-Quartile Men vs. Support Ratio, by County



Notes: This figure shows correlations between our social capital measures and mean life expectancy at age 40 for men with income in the bottom income quartile across counties. Data on life expectancy is obtained from publicly available data released by Chetty et al. (2016c), who construct these estimates using information from tax records linked to death certificate data for the U.S. population. Panel A replicates Figure 4a, using life expectancy for low-income men as the outcome instead of upward mobility, and weighting by the number of bottom-income-quartile men in each county. Panels B and C display binned scatter plots of life expectancy for low-income men vs. mean clustering coefficients and support ratios by county, respectively, again weighting by the number of bottom-income-quartile men in each county. See notes to Figure 4 and Figure 11 for details on the construction of these figures.

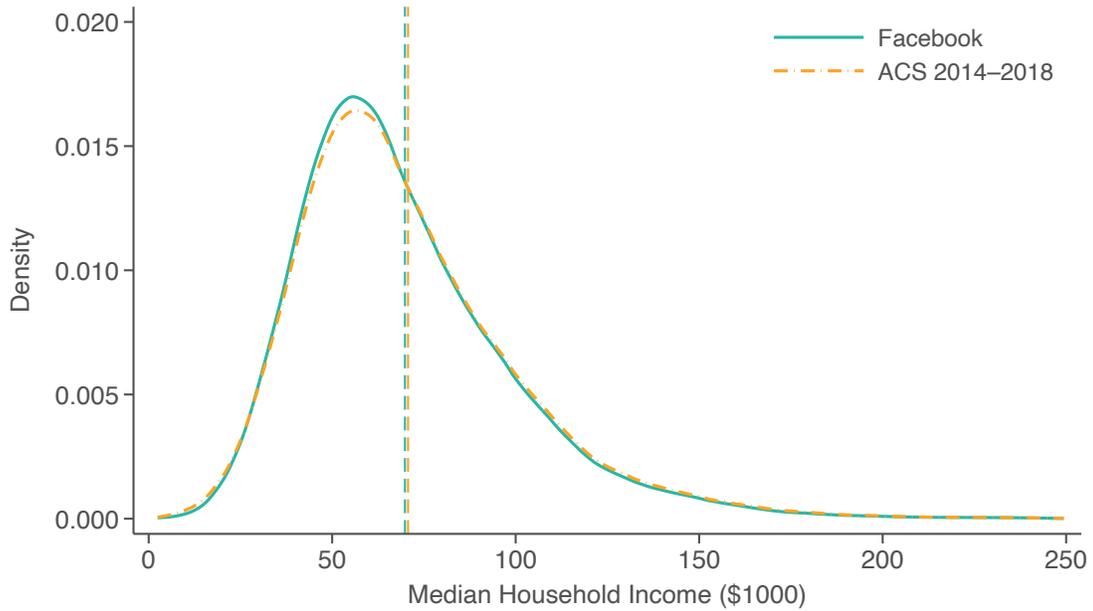
SUPPLEMENTARY FIGURE 12: Relative Geographic Coverage of Facebook Data



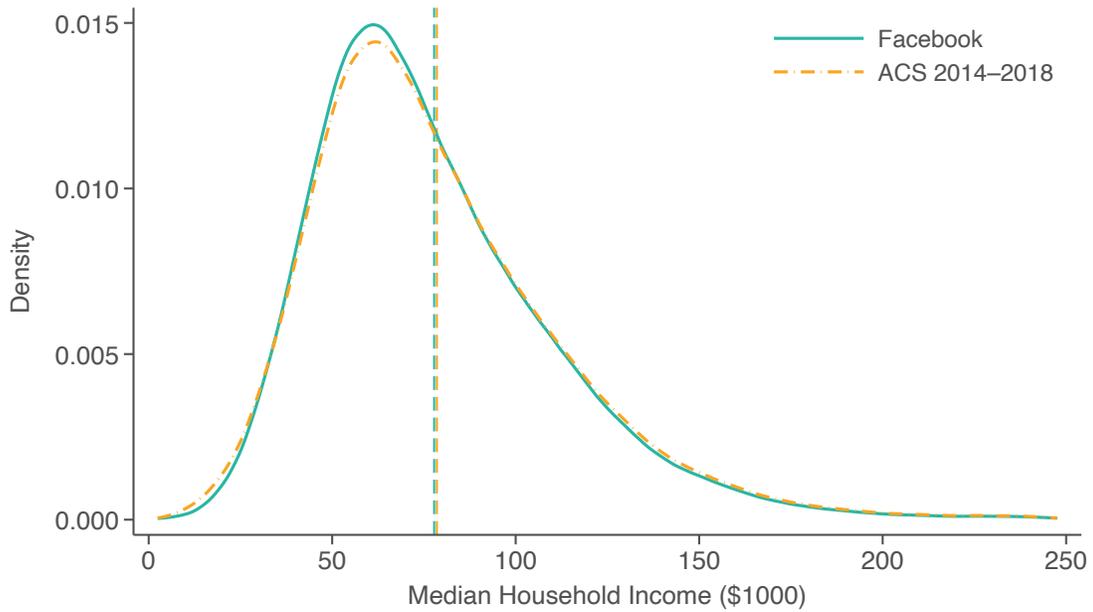
Notes: This figure shows the relative geographic variation in Facebook coverage across counties for our primary analysis sample. For each county, we compute the share of the Facebook users in our data relative to the number of 25–44 year-olds in the 2014–2018 American Community Survey (ACS). We normalize these county-level coverage estimates by the corresponding U.S.-wide coverage in our sample to obtain relative coverage estimates. For example, a value of 1.2 for a given county implies that our sample coverage rate is 20% higher in that county relative to the overall national sample coverage rate.

SUPPLEMENTARY FIGURE 13: Distribution of ZIP Code-Level Incomes in Facebook Data vs. ACS

A. Age 25 to 44

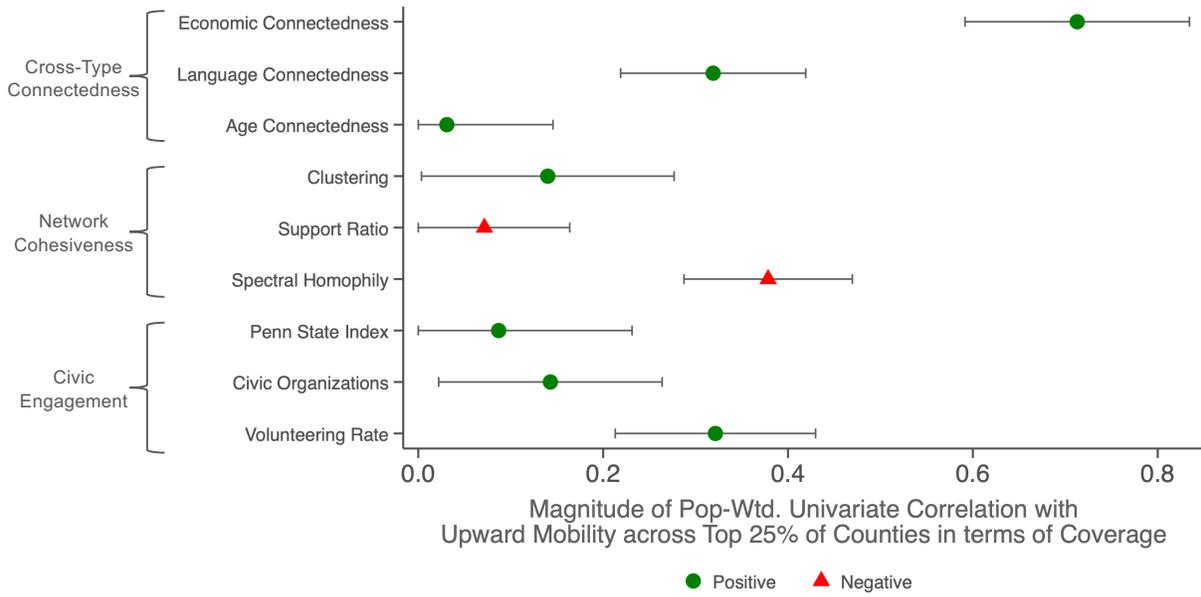


B. Age 45 to 64

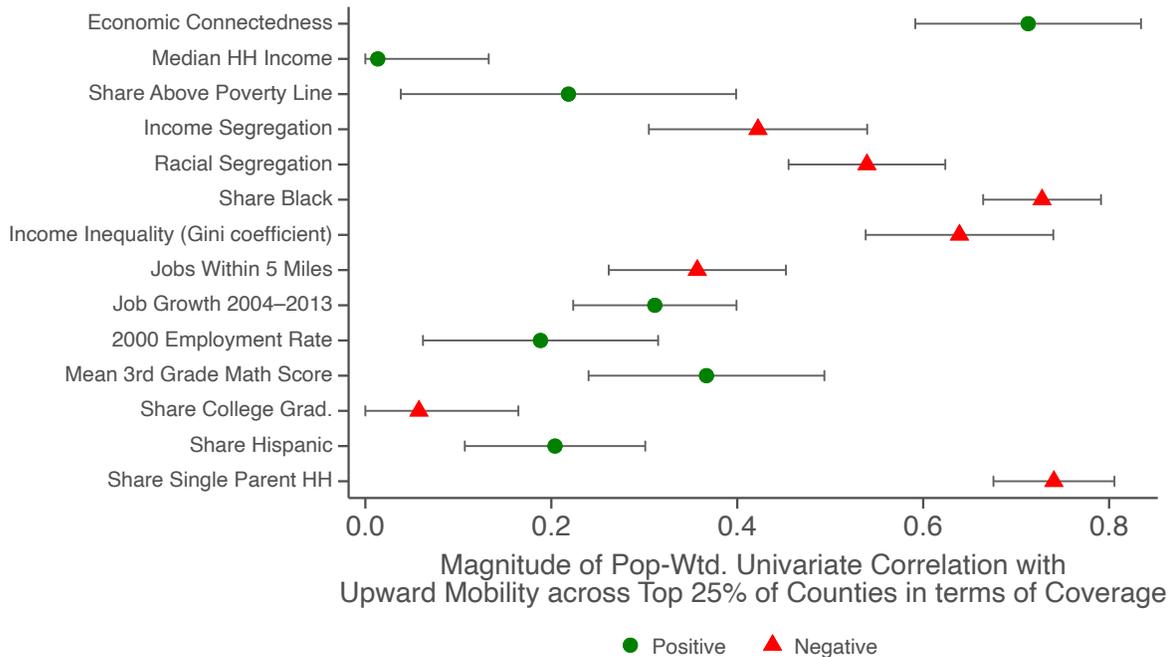


Notes: This figure evaluates the geographic representativeness of the Facebook sample in terms of median incomes by ZIP code for the 25 to 44 (Panel A) and 45 to 64 (Panel B) age groups. The dashed orange curve shows the distribution of median household income by ZIP code (as measured in the 2014–2018 American Community Survey (ACS)), weighting by the number of individuals in the relevant age group residing in each ZIP code according to the ACS. The solid green curve replicates the same density plot of ACS median household income by ZIP code, weighting instead by the number of individuals residing in each ZIP code in the relevant age group in the Facebook data. The dashed vertical lines show the median incomes in each sample (green for Facebook, orange for ACS).

SUPPLEMENTARY FIGURE 14: County-Level Correlations for Top 25% of Counties by Facebook Coverage Rates

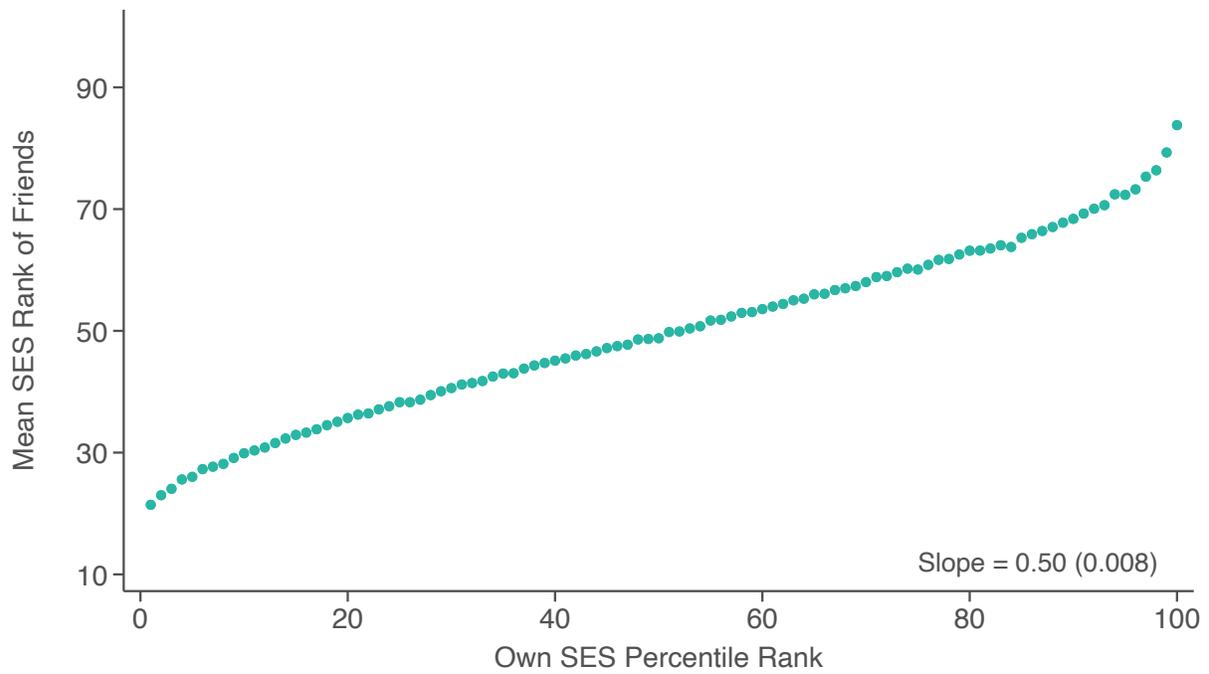


B. Upward Mobility vs. Neighbourhood Characteristics



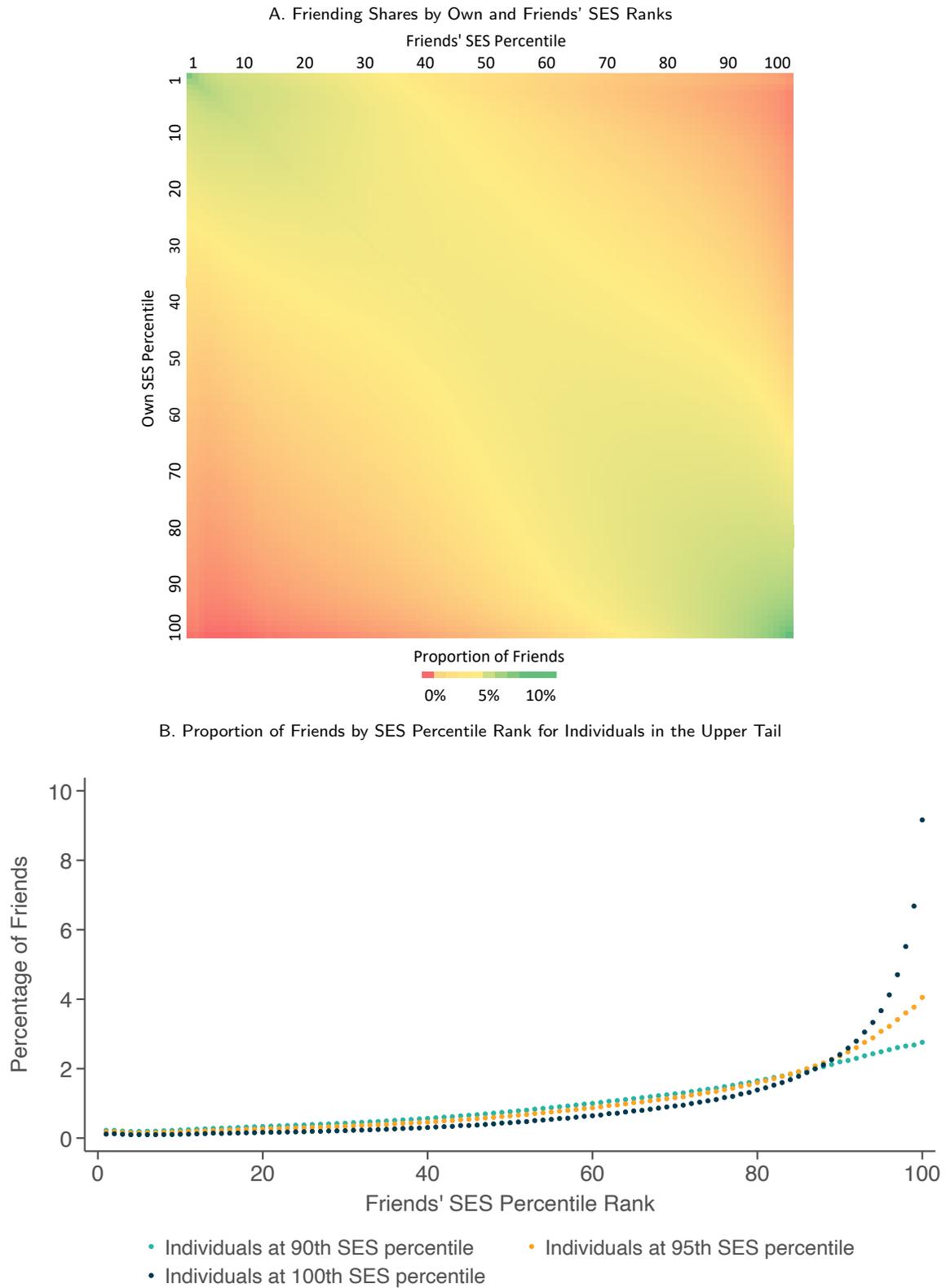
Notes: Panel A replicates Panel A of Figure 4, restricting the sample to the top 25% of counties (unweighted) in terms of Facebook coverage. Coverage is defined as the number of individuals in our primary analysis sample divided by the number of 25–44 year-olds in the 2014–2018 American Community Survey (ACS) in each county. Panel B replicates Panel A of Figure 10, again restricting the sample to the top 25% of counties in terms of coverage.

SUPPLEMENTARY FIGURE 15: Relationship Between Friends' and Own Socioeconomic Status on Instagram



Notes: This figure replicates the series in circles in Figure 1a, plotting the mean SES rank of individuals' friends vs. individuals' own SES ranks in the Instagram sample. See Supplementary Information A.4 for details on the construction of the Instagram sample and definition of SES ranks in that sample.

SUPPLEMENTARY FIGURE 16: Proportion of Friendships by SES Percentile Rank



Notes: Panel A presents a heat map showing the proportion of friends in each SES percentile rank by an individual's own SES percentile rank. Red squares represent lower proportions of friends in a given SES percentile, whereas green squares represent higher proportions of friends in that percentile. Panel A replicates Table 1a at the percentile level instead of the decile level. Panel B shows the proportion of friends by friends' SES percentile rank for individuals at the 90th, 95th, and 100th percentiles of the national SES distribution.