SOCIAL CONNECTEDNESS: MEASUREMENT, DETERMINANTS, AND EFFECTS

ONLINE APPENDIX *

Mike Bailey† Ruiqing Cao‡ Theresa Kuchler§ Johannes Stroebel¶ Arlene Wong||

In this Appendix, we provide further details on the empirical analyses conducted in the main body of the paper. We also provide more detailed descriptions of the data used in the main analyses, and present additional explorations of the empirical patterns. The Appendix is structured along the same sections as the main body of the paper.

A The Determinants of Cross-County Social Connectedness

In the main body of the paper, we explored a number of determinants of the intensity of social connectedness between US counties. We next provide additional details on the empirical patterns documented there.

In Figure 1 of the main body, we plotted heat maps of the "relative probability of friendship" between two counties (San Francisco County and Kern County) and the rest of the United States. In this Appendix, we show similar heat maps for a number of other US counties. The "relative probability of friendship" measure is constructed by dividing the Social Connections Index between counties i and j by the product of the number of Facebook users in the two counties:

$$RelativeProbFriendship_{i,j} = \frac{SCI_{i,j}}{FB\_Users_i \times FB\_Users_j}. \tag{A1}$$

Dividing by the product of the number users takes into account that we will see more friendship links between counties with more Facebook users. Only relative magnitudes of $RelativeProbFriendship_{i,j}$ can be interpreted: if it is twice as large, a given Facebook user in county i is twice as likely to be connected to a given Facebook user in county j.

Share of Friends within Geographic Distances. Table 1 in the main body of the paper presented summary statistics on the geographic concentration of US friendship networks. Appendix Table A1 provides additional information. Columns 1-3 and columns 5-7 summarize to the share of friends and population living within 50, 100 and 200 miles of each county, respectively. These columns are

*This document contains the Online Appendix to "Social Connectedness: Measurement, Determinants, and Effects" by Mike Bailey, Ruiqing Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, published in the Spring 2018 issue of the *Journal of Economic Perspectives*.

†Facebook. Email: mcbailey@fb.com

‡Harvard University. Email: ruiqingcao@fas.harvard.edu

§New York University, Stern School of Business. Email: tkuchler@stern.nyu.edu

¶New York University, Stern School of Business, NBER, and CEPR. Email: johannes.stroebel@nyu.edu

|| Princeton University, and NBER. Email: arlenewong@princeton.edu

the same as Table 1 in the main body of the paper. We also include information on the share of friends living within 500 miles of a given county (column 4) and and the share of the US population living within 500 miles of a given county (column 8). We see that for the average county, over 79% of friends live within 500 miles, despite the fact that, for the average county, only 39.7% of the population live within 500 miles. This is consistent with the discussion in the main body of the paper, where we highlighted that US friendship networks are substantially more concentrated than the population distribution.

We next analyze the relationship between geographic distance and friendship links across county-pairs more systematically. An existing literature has suggested that the relationship between the probability of friendship between any two individuals, $P(d)$, and the geographic distance between the two individuals, d, can be represented by the relationship $P(d) \sim d^{\alpha}$. The estimates for the parameter α, which captures the elasticity of friendship probability with respect to geographic distance, vary significantly across settings, including estimates of -2 in a study of cell phone communication networks in the United Kingdom Lambiotte et al. (2008), estimates of -1 among bloggers Liben-Nowell et al. (2005), and estimates of -0.5 in location-based online social networks such as Brightkite, Foursquare, and Gowalla Scellato et al. (2011).

In Appendix Figure A1, we plot a binned scatter plot analyzing the relationship between the log of geographic distance on the horizontal axis and the log of the SCI on the vertical axis. In the construction of this graph, we control flexibly for the log of the product of the two counties' populations by including 50 dummy variables for equal-sized quantiles of the distribution. Two counties have more friendship links when they are closer geographically, but the elasticities are more negative at shorter geographic distances between the two counties.

To obtain magnitudes for the associated elasticities, we estimate regression A2. The unit of observation is a county-pair. The dependent variable is the log of the SCI, while $\log(pop_i \times pop_j)$ denotes the log of the product of the county-populations and $\log(d_{ij})$ denotes the log of the geographic distance between i and j.

$$\log(SCI_{ij}) = \beta_0 + \beta_1 \log(pop_i \times pop_j) + \beta_2 \log(d_{ij}) + \epsilon_{ij} \tag{A2}$$

Appendix Table A2 presents the estimated results from regression A2. Column 1 presents estimates of β_1 when we do not also control for $\log(d_{ij})$. The elasticity of social connectedness with respect to the product of the county populations is slightly larger than one. Overall, the differences in the populations can explain about 68% of the variation in the number of friendship links across counties. In column 2, we also control for the log of geographic distance. Over the entire range of distances, the average estimated elasticity between geographic distance and friendship links is about -1.07. The addition of this further control variable increases the R^2 of the regression to 81%. This suggests that geographic distance is able to explain a significant amount of the cross-county-pair variation in social connectedness. For comparison, columns 3 to 7 include the five columns of Table 2 from the main body of the paper.

In column 5 of Table 2 in the main body of the paper (which corresponds to column 7 in Appendix Table A2), we control for differences in socioeconomic characteristics across counties in the regression explaining the social connectedness between county pairs (equation A2). We next discuss the data sources for these socioeconomic measures, and describe the magnitudes of the detected homophily.

Data Sources. Data on income, racial composition, and education levels come from the 5-year estimates of the 2013 American Community Survey. County-level voting data for the 2008 presidential election was provided by The Guardian (2009). The major religious traditions we consider are Evangelical Protestant, Mainline Protestant, Historically Black Protestant, Roman Catholic, Jewish, Latter-day Saints (Mormon), Islamic, Hindu, Buddhist, Orthodox Christian, and Jehovah's Witnesses. Data are collected by Infogroup (2009) based on its database of more than 350,000 houses of worship.

Magnitudes. The inclusion of additional demographic differences between states has little effect on the estimated elasticity of friendship links with respect to geographic distance. The estimates on the controls for demographic differences suggest that a $10,000 (0.69 standard deviation) increase in the difference of mean incomes between two counties is associated with a 6% decline in the number of friendship links between these counties. Similarly, a ten percentage point (1.9 standard deviation) increase in the difference in the share of population without a high school degree is associated with a 12% decline in the number of friendship links. A 10 percentage point (0.88 standard deviation) increase in the difference in the share of votes for Obama in 2008 is associated with a 6% decline in friendship links. And lastly, a 10 percentage point (0.75 standard deviation) increase in the difference in the number of religious congregation members is associated with a 2% decline in the number of friendship links. However, despite the statistical and economic significance of these effects, the increase in the R^2 between columns 2 and 5 of Table 2 is relatively modest. This suggests that, relative to geographic distance, differences in socioeconomic characteristics explain significantly less of the cross-county-pair variation in social connectedness.[1]

State-State Adjacency Matrix. Appendix Figure A2 describes the strength of friendship links between US states. This adjacency matrix plots the percentile rank of the relative probability of a friendship link between a Facebook user in state i and a Facebook user in state j. This relative probability is constructed similarly to equation A1 above, by taking the total number of friendship links (i.e., the SCI) between each pair of states, and dividing this by the product of the number of Facebook users in both states. Darker colors correspond to states that are more strongly connected. States are organized by US Census Bureau Divisions. There are strong connections within census divisions, as well as between geographically adjacent divisions (which may not be adjacent by division number). Washington, D.C., is very well-connected to most states in the United States, regardless of geographic distance. Other strong connections between geographically dispersed regions are potentially explained by migration or tourism. For example, both Colorado and Hawaii are well-connected to many different states across the United States.

Connected Communities. In the main body of the paper, we presented information on the communities that would form if we grouped together US counties to create 20 communities with the aim of maximizing within-community social connectedness. We next provide additional details on the ag-

[1] Appendix Figure A3 shows binned scatter plots at the county-pair level that portray the univariate relationship between differences across the counties along these socioeconomic measures and the SCI between the counties. These plots are consistent with the multivariate regression results in Table 2, and show that the SCI between two counties is lower if the difference between the two counties on any of the given socioeconomic indicators increases. All of these plots control flexibly for the log of the product of the counties' populations and the log of the distance between each pair of counties.

glomorative clustering algorithm that we used to construct the communities, before discussing the communities formed when we split the United States into 50 or 75 units.

Algorithm. Conceptually, the agglomorative clustering algorithm starts by considering each of the N counties in the US as a separate community of size one. In the first step, the two "closest" counties are merged into one larger community, producing N-1 total communities. In each subsequent step, the closest two communities are again merged. This process continues until all the counties are merged into a given number of clusters. We define the "distance" between two counties as the inverse of $RelativeProbFriendship_{i,j}$ in equation A1: the lower the probability of a given Facebook user in county i knowing a given Facebook user in county j, the "farther apart" socially the two counties are. We calculate the closeness between communities with more than one county as the average distance between the counties in the communities.

50 Units. In Appendix Figure A4, we use this algorithm to group the United States into 50 distinct communities. Many multi-state groups from Figure 2 in the main body of the paper are now split into separate communities for each state. In addition, many states are now split into separate communities. California divides into a region around Los Angeles, a region around San Diego, and the rest of the state; the most northern California counties form a community with Oregon and Washington state. Texas is further divided into North and South Texas, and Southern Florida is separated from a northern part that is joined with the region around Savannah, Georgia. Philadelphia and New York City form communities that are separate from the rest of Pennsylvania and New York State, respectively.

75 Units. In Appendix Figure A5, we group the United States into 75 distinct communities, creating additional sub-communities within states. Many states group into eastern and western communities, like Virgina, Pennsylvania, Ohio, and Kentucky. Other regions separate into northern and southern communities, as seen in the division of Illinois, the dissolution of groupings like the Carolinas and the Pacific Northwest states, and New England's splintering into two groupings of three states each. The Appalachian region breaks into more small communities as a new cluster emerges in eastern Tennessee, eastern Kentucky and western Virginia separate from the rest of their states, and Western Maryland and West Virginia's eastern panhandle also join together. Florida, previously divided into a northern and southern portion, is now five distinct communities as the southern portion breaks into quarters. The large states of California and Texas, already grouped into a number of different communities, divide further. In California, the Bay Area and the region north of Los Angeles each break away from the large central region seen in Appendix Figure A4. Texas, meanwhile, adds an eastern division that includes both of its two largest cities, Houston and Dallas, and a triangular grouping beneath the Texas panhandle and Oklahoma also emerges.

Effect of State and Regional Borders. We also further explore the role of state and regional borders in shaping social connectedness. In particular, Table 2 and Figure 2 in the main text, as well as Appendix Figures A4 and A5, highlight that social connectedness is significantly stronger within states than it is across state lines.

More evidence for the important role of state borders is provided by the friendship networks plotted in Appendix Figure A6, which shows the distribution of the social networks of a number of US counties. In each of the panels, the friendship networks are most dense within the state in which the county is located, and the probability of a friendship link diminishes once state lines are crossed. In Panel A, the friendship network of Macomb County, MI, shows strong connections to both Michigan's upper and lower peninsulas, and less strong connections to counties across the state borders with Indiana and Wisconsin. This example highlights that the state-border effect is not specific to counties in the center of the state, or to small counties that may have a limited number of total connections. Indeed, Macomb is the third-largest county in Michigan and neighbors Wayne County, home to Detroit, in southeastern Michigan. Panels B, C, and D of Appendix Figure A6 show the friendship networks of Erie County, NY, Bexar County, TX, and Schuylkill County, PA, respectively. All of these counties display similarly strong state-border effects on the geographic distribution of friendship links. Panel E shows the friendship network of Marion County, KY. Of the ten counties in the United States with the highest share of friends within 100 miles, seven are in Kentucky, and of the 25 counties with the highest share of friends within 100 miles, 19 are in Kentucky. Unsurprisingly, plotting the relative probability of friendship links of counties in Kentucky reveals many examples of friendship networks that are very dense within the state. Panel F shows the friendship network of Clark County, IN, which is on the border with Kentucky. For this border county, state-border effects for both Kentucky and Indiana are strongly pronounced.

In addition to the state-border effects documented above, there are some groupings of states that show a high degree of mutual connectivity. In Appendix Figure A4, when counties were divided into 50 clusters based on their connectivity, state borders were mostly preserved. Notable exceptions were across-state groups that were formed by the six New England states and by North and South Carolina (see also Appendix Figure A2). Panel G of Appendix Figure A6 shows the friendship networks of Bristol County, MA, revealing strong connections with the entire New England region; the border effect manifests itself outside this region. Likewise, Panel H shows the friendship network of Allendale County, SC, to all other counties in the continental United States. We see strong connections to counties in North and South Carolina, and a decline at the borders of this region.

To better understand which counties display border effects, we run regression A3 separately for each county i. The unit of observation is a county-pair. The dependent variable is the log of the SCI. $\mathbb{1}_{\text{Same State}}$ is an indicator variable that is set equal to one if the two counties are in the same state; and $g(d_{ij})$ flexibly controls for the geographic distance between i and j. In our baseline specification, this is achieved by grouping county-pairs into 250 bins by the distance between them, and then including separate indicator variables for each group.

$$\log(SCI_{ij}) = \beta_0 + \beta_1 * \mathbb{1}_{\text{Same State}} + \beta_2 g(d_{ij}) + \epsilon_{ij} \tag{A3}$$

Appendix Figure A7 maps the coefficient β_1 for each county in the continental United States. Red counties reflect counties with stronger state-border effects, with higher values for β_1, while blue regions exhibit weaker state-border effects. There is strong heterogeneity in the distribution of state-border effects across states. Some states display strong state-border effects across almost all counties, while others do not. Most states display a mixture of regions with high state-border effects, typically located in the more central regions of the states, along with areas with lower state-border effects along their borders with other states.

To examine the characteristics of counties with strong state-border effects, Appendix Figure A8 shows county-level binned scatter plots of coefficient β_1 and the share of friends within 100 miles in three specifications: Panel A does not include fixed effects, Panel B controls for state fixed effects, and Panel C controls for commuting zone fixed effects. All three panels display a roughly parabolic relationship, with counties that have a low state-border effect generally having a low share of friends within 100 miles, with the share of friends within 100 miles rising as the state-border effect increases before falling for counties with a high state-border effect. However, none of the relationships are particularly strong. Appendix Figures A9, A10, and A11 show the calculated state-border effect plotted against a number of county-level measures of socioeconomic outcomes, much like the plots of the share of friends within 100 miles against the same socioeconomic outcomes in Figure 3 in the main body of the paper. Within a state, richer counties show weaker state-border effects, while counties with higher measures of social capital show stronger state-border effects. The state-border effect does not appear to be strongly correlated with the other outcome variables studied in these figures.

Effect of Physical and Topological Geographic Features. Physical barriers to connectivity may help explain some of the patterns observed in our discussion of state-border effects and the variation in the geographic extent of friendship networks. In many cases, state borders are partly determined by geographic features, such as the borders of states following the Mississippi River or the Appalachian Mountains. Appendix Figure A12 shows two examples of geographic features exerting a strong influence on the geographic spread of friendship networks. Panel A displays the relative probability of friendship links to Scott County, AR. The friendship network of this county is significantly weaker once the Mississippi River is crossed. However, this is hard to separate from the state-border effect. Panel B plots the relative probability of friendship links to Belmont County, OH. There are strong friendship links within Pennsylvania up until the Appalachian Mountains, and linkages are also strong through West Virginia until the border with Virgina, marked by the Blue Ridge Mountains (see Wikimedia, 2010; Encyclopedia Britannica, 2012, for maps of these mountain ranges). This demonstrates that evidence of the potential physical determination of friendship networks is not limited to instances that may also involve state borders. This observation is highly consistent with the findings in Appendix Figures A4 and A5, which showed that our clustering algorithm splits the central and south-central Appalachian region into a relatively large number of small distinct communities. Mountain regions, historically, have been home to many isolated, often culturally and linguistically distinct, populations due to their inaccessibility, and this is still true of the Appalachian regions today (see Dial (1969) for more information on Appalachian dialects and New York Times (2008) for a discussion of the linguistic diversity of the similarly mountainous Caucasus region).

Further Explorations of Within-US Social Connectedness. While we have previously documented a number of strong patterns in social connectedness across US counties (i.e., it declines in geographic distance, and at state borders and physical barriers), social networks differ significantly across counties in the same geographic regions. The SCI data enables us to highlight a number of interesting patterns, allowing us to document the role that heterogeneity in the demographics, histories, and industrial compositions of counties plays in shaping social networks.

Figure A13 shows the social networks of three different Illinois counties. Panel A plots the relative probability of friendship links to McHenry County, IL, home to some of the northern suburbs of Chicago. Counties with a high probability of connection to McHenry County are generally distributed throughout the upper Midwest and include all of Illinois. Further, since McHenry County lies along the border with Wisconsin, it displays a high probability of connection to counties across the entirety of that state. There is also a pocket of strong connectivity to Colorado, perhaps revealing an affinity for winter sports among McHenry County's generally upper-middle-class population. In Panel B, the probability of friendship links to Cook County, IL, is plotted. Cook County is home to Chicago proper, and differs from suburban McHenry County along several demographic dimensions. As a result, the geographic spread of Cook County's friendship network looks radically different. Indeed, Cook County's friendship links show strong connections to the South. This pattern is consistent with the mass migration of southern African Americans to northern and Midwestern cities throughout the twentieth century. This movement from south to north, known as the "Great Migration," resulted in over four million southern-born African Americans living outside of the South by 1980 (see Crew, 1987; Tolnay, 2003, for more information). Many of these migrants moved to large northern and Midwestern cities, like Chicago and Milwaukee. Panel C of Figure A13 shows the distribution of the relative probability of friendship links for Crawford County, IL, which also has a high concentration of links to Louisiana and Mississippi. Yet, the large migration that likely contributed to shaping Cook County's friendship network did not cause a demographic transformation for this mostly rural county, which does not have a large African-American population. One potential explanation for the pattern exhibited here is the industrial composition of the county. The largest city in the county, Robinson, is home to a large oil refinery, and Crawford County's connections in the South are primarily focused along the oil-producing Gulf Coast and in Texas. Indeed, Crawford County's oil refinery employs over 1,000 workers in a county with under 20,000 total inhabitants (see Marathon Petroleum, 2016, for more information). Other oil-producing counties, such as McKenzie County, ND, exhibit similar patterns of social connectedness (see Appendix Figure A17, and the associated discussion below).

Counties within Wisconsin also display significant heterogeneity in the geographic distribution of their friendship networks. Panel A of Appendix Figure A14 maps the distribution of friendship links to Manitowoc County, WI, which are strongest in the upper Midwest. In Panel B, the plot of the relative probability of friendship links to Milwaukee County, WI, shows strong connections to counties in the southern United States. As in the case of Cook County and Chicago, this is likely a result of the Great Migration-era movement of African Americans from the South to Milwaukee. Panel C shows the friendship network of Menominee County, WI. This map reveals a high degree of connectivity to counties in the West and in Oklahoma. Menominee County is coterminous with the Menominee Indian Reservation, and the counties that have a high probability of connectivity to Menominee County generally have large populations of Native Americans or are home to reservations (see Amauta, 2010; National Park Service, 2003, for maps showing the distribution of Native American populations and locations of reservations, respectively).

These figures reveal that large population movements can have lasting effects on the geographic distribution of social networks. This is highlighted by the persistence of friendship links to the South for counties that experienced a large inflow of migrants during the Great Migration. The same patterns were also revealed when analyzing the friendship links of Kern County, CA, to Oklahoma and Arkansas, the origin counties of the Dust Bowl migrants in the 1930s.

While past population flows are a key determinant of present-day social connectedness, the impact of ongoing population flows can be even more apparent. Panel A of Appendix Figure A15 shows the geographic distribution of the friendship networks of Rapides Parish, LA. Counties with the highest probability of connection to Rapides Parish are primarily in the South. In contrast, Panel B shows the friendship networks of the neighboring parish, Vernon Parish, LA. Vernon Parish shows very high levels of social connectivity to a much greater swath of the South, stretching into the Midwest. There are also strong connections throughout much of the United States. The presence of a large army installation in Vernon Parish likely explains the difference in the social networks of these neighboring parishes. Vernon Parish is home to Fort Polk, and troops stationed at the base make up roughly a fifth of the county's population, while Rapides Parish has no military presence.

Panel A of Appendix Figure A16 plots the friendship network of Miami-Dade County, FL. Counties with a high probability of connection to Miami-Dade are mostly within Florida, with some counties in the New York area also showing a high degree of connectedness to Miami. In Panel B, the friendship network of Charlotte County, FL, is mapped. Here, there is a high probability of connection to much of the Midwest and the Northeast, particularly to Michigan and New England. The over-65 share of the population in Charlotte County is the highest in the country, at 34.1%, and the median age of Charlotte County is 54.3 years. Charlotte County's unique demographics are due to its popularity as a retirement destination, which also explains its strong connections to the Midwest and Northeast. Panels C and D show the friendship networks of Collier County, FL, and Palm Beach County, FL. Collier County, on the western coast of Florida, shows stronger connectivity to the Midwest than does Palm Beach, on the eastern coast of the state. Both counties have similar demographics, which suggests that the differences between the spread of their friendship connections is potentially related to their relative geographic proximity to the Midwest and Northeast, respectively, and their popularity as a destination for tourism or retirement.

Recent advances in horizontal drilling and hydraulic fracturing ("fracking") have enabled a dramatic expansion in oil production in states that had not previously been large oil producers. The Bakken formation in western North Dakota has proven to be particularly productive. Panel A of Appendix Figure A17 shows the friendship network of Richlands County, ND, which is located on the eastern border of the state away from the Bakken formation. Panel B shows the plot for McKenzie County, ND, which is located along the western border of the state in the Bakken formation. McKenzie County has rapidly become one of the most productive oil-producing counties in the United States. The influx of oil workers from across the United States, particularly from other states in the West as well as oil-producing regions in Texas and along the Gulf Coast, results in a high connectedness to most counties in the country (see NPR, 2015, for more information).

Panel A of Appendix Figure A18 shows the friendship network of Sanpete County, UT, a primarily rural county with high probabilities of connection across the Mountain States. Panel B shows the friendship network of Summit County, UT, which contains many winter sports retreats, including the resorts that hosted skiing and snowboarding events at the 2002 Winter Olympics. The distribution of friendship links across the western United States is essentially the same for these two counties, but Summit County also shows a high probability of connection to counties in New England, many of which are also winter sport destinations.

B Concentration of Social Networks and County Characteristics

In the main body of the paper, we explored the relationship between the geographic concentration of a county's social network, measured as the share of friends living within 100 miles, and socio-economic outcomes at the county level.

Additional Details on Figure 3. We next discuss the magnitudes and potential theoretical underpinnings of the relationships uncovered in Figure 3 in the main body of the paper. In Appendix Figure A19, we reproduce Figure 3. In addition, Appendix Figure A19 also includes information on additional socio-economic outcomes, including labor force participation, casual social mobility, and life expectancy conditional on race.

Panel A of Appendix Figure A19 shows that counties with higher average income have more dispersed friendship networks. The relationship is not linear: mean household incomes are roughly flat at $60,000 to $65,000 for counties that have a share of friends living within 100 miles between 30% and 65%. Once the share of friends living within 100 miles exceeds 65%, mean household incomes drop substantially, eventually falling below $50,000. Panel B suggests that counties with more geographically dispersed friendship networks have higher labor force participation rates, again with a strong relationship among counties with more than 65% of friends living within 100 miles. Panel C documents that counties with more concentrated friendship networks have lower education levels, measured by the share of county population without a high school degree. Panel D shows that counties with more concentrated friendship networks have higher rates of teen pregnancy.

Panels E and F of Appendix Figure A19 correlate the geographic dispersion of social networks at the county level with measures of social mobility. In Panel E we use the measure of absolute social mobility from Chetty et al. (2014). This measure captures the expected rank in the national income distribution at adulthood of children whose parents are at the 25th percentile of the national income distribution. In Panel F we use estimates of the causal effect on social mobility from living in the county from Chetty and Hendren (2015). This effect is measured as the percentage gains (or losses) in income at age 26 relative to the national mean from spending one more year in the county for a person at the 25th percentile in the national income distribution. Those counties with more geographically concentrated friendship networks have lower social mobility on both measures. This relationship appears across the entire range of the geographic concentration of social networks.

In Panel G of Appendix Figure A19, we show the correlation between the share of friends living within 100 miles, and a measure of social capital from Rupasingha, Goetz and Freshwater (2006). This measure of social capital aims to capture the intensity of social interactions at the local community level, and uses a number of input variables, including voter turnout rates, the fraction of people who return their census forms, and different measures of participation in community organizations. Counties with a higher social capital index have less geographically concentrated social networks. This suggests that being more actively involved in local communities does not come at the expense of having a more geographically concentrated social network. Instead, these results suggest that those counties that see more active community engagement also have social networks with a broader geographic reach.

A large literature has analyzed the relationship between social interactions and health outcomes, with much research concluding that there is a causal positive effect of social relationships on health (see the literature reviews in House et al., 1988; Holt-Lunstad, Smith and Layton, 2010). To test whether a correlation between the geographic concentration of social networks and life expectancy is also present in the SCI data, in Panels H and I of Appendix Figure A19 we consider data on the life expectancy of a male at the first quartile of the national income distribution, as reported by Chetty et al. (2016). In Panel H we analyze the unconditional life expectancy, in Panel I the life expectancy conditional on race. Across both measures, more geographically concentrated social networks are associated with shorter life expectancy.

Conditioning on State or Commuting Zone. Appendix Figures A20 and A21 show similar relationships as in Figure 3 of the main body and Appendix Figure A19, but also condition on the state and the commuting zone, respectively. We use the commuting zone definitions based on commuting patterns in the 1990 Census constructed by Tolbert and Sizer (1996). Commuting zones are designed to span the area in which people live and work. Including state and commuting zone fixed effects allows us to compare counties that are geographically close to each other, which ensures that our results are not driven by differences in population density, which might affect the number of people living within 100 miles. Most of these plots show patterns very similar to those in Figure 3 and Appendix Figure A19. The notable exception is Panel F in both figures, showing causal social mobility as defined by Chetty and Hendren (2015). When fixed effects for state and commuting zone are excluded, as in Appendix Figure A19, the plot shows causal social mobility to decline as the share of friends within 100 miles increases; including fixed effects for state or commuting zone causes this relationship to reverse, with causal social mobility rising as the share of friends within 100 miles increases.

Alternative Density Measure. We also explore a second measure of the density of social networks, namely the share of friends among the nearest 50 million people in and surrounding a county. Appendix Figure A22 shows a heatmap of the density of frienship links using this measure. The Northeast and portions of the Midwest display less densely concentrated friendship networks while portions of the South, Plains, and Mountain States exhibit more dense social networks compared to our previous measure. Differences in the two measures of concentration are the result of variation in population density across the United States.

Appendix Figures A23, A24, and A25 present county-level binned scatter plots of the share of friends living among the nearest 50 million people by demographic characteristics. Figures A24 and A25 are conditional on state and commuting zone, respectively. For certain demographic characteristics, most notably average income (Panel A), labor force participation (Panel B), share with no high school degree (Panel C), and the teenage birth rate (Panel D), a stronger correlation is apparent for this measure of the density of social networks than for the share of friends within 100 miles. Other relationships, in particular absolute social mobility, causal social mobility, and social capital show weaker correlations in Appendix Figures A23, A24, and A25 than the corresponding Panels in Figure 3 and Appendix Figures A19, A20 and A21. For some of the demographic characteristics, the geographic concentration of social networks shown in Figure 3 and Appendix Figure A20 has greater predictive power while for others the density of social networks as shown in Appendix Figure A23 is a stronger predictor. The R^2 of the quadratic regressions that underlie each of the Panels in Appendix Figure A20

[Appendix Figure A23] are: 8.7% [28.4%] for average income (Panel A), 4.1% [14.7%] for labor force participation (Panel B), 15.8% [24.5%] for share with no high school degree (Panel C), 6.2% [18.2%] for the teenage birth rate (Panel D), 5.7% [0.2%] for absolute social mobility (Panel E), 3.5% [0.4%] for causal social mobility (Panel F), 12.2% [0.8%] for social capital (Panel G), 13.5% [10.5%] for male life expectancy (Panel H), and 10.3% [9.3%] for male life expectancy conditional on race (Panel I). In particular, Panels A, B, C, and D exhibit significantly stronger correlations with the density of social networks rather than the geographic concentration, while relationships are weaker for the density of social networks in all other panels.

Multivariate Analysis. The previous analyses have explored univariate correlations between measures of the concentration of social networks and outcome variables of interest. However, many of these outcome variables are potentially correlated. We therefore also conduct a multivariate analysis between our measures of geographic concentration and socioeconomic outcomes at the county level. In columns 1 to 3 of Appendix Table A3, we analyze the correlation with the share of friends living within 100 miles, and in columns 4 to 6 with the share of friends within the closest 50 million people. In columns 2 and 5, we also control for state fixed effects, and in columns 3 and 6 we also control for commuting zone fixed effects. We do not include all of the nine different outcome measures studied above, since many of them are highly collinear: our final specification controls for average income (which is highly correlated with educational outcomes), causal social mobility, social capital, and one of the two life-expectancy measures. All of the univariate relationships are recovered in this multivariate analysis. The one important change is that, once we control for other socioeconomic outcomes, causal social mobility is always higher in areas with more concentrated social networks, whether or not we condition on state or commuting zone fixed effects. Understanding this potentially counter-intuitive relationship is an exciting area for further research.

C Social Connectedness and Cross-County Activity

In the main body of the paper, we analyze how across-region social connectedness is related to three measures of across-region activity: trade, patent citations, and migration. In this Appendix, we provide further details on each of these analyses.

C.1 Social Connectedness and Within-US Trade Flows

Data Description. In the main body of the paper, we use US state-level trade flows data from the Commodity Flow Survey (CFS) to measure interstate trading volumes. These data are collected through a survey of establishments by the US Census Bureau every five years. We follow Yilmazkuday (2012), and exclude observations that have not been disclosed by the Census because of high coefficients of variation (greater than 50 percent). These observations are marked with an "S" in the data.

Regression Analysis. Table 3 of the main body of the paper describes the correlation between trade flows and social connectedness. The estimates were based on the “gravity equation” given by regression A4:

$$\log(v_{ij}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \epsilon_{ij}. \tag{A4}$$

The dependent variable, $\log(v_{ij})$, captures the log of the value of trade in 2012 between origination state i and destination state j. The variable $\log(d_{ij})$ denotes the log of geographic distance between states i and j,[2] and the variable $\log(f_{ij})$ denotes the log of the relative number of friendship links between the states (i.e., the log of the SCI). We include fixed effects for each state, denoted by ψ_i and ψ_j, which capture state-specific characteristics. We also include dummy variables for own-state flows, and dummy variables if the states are adjacent to each other, to control for factors affecting trade flows across borders. In some specifications, we include other control variables, given by X_{ij}, to capture differences between states i and j on measures such as GDP per capita, unemployment rates, sectoral composition, union share, and population density.[3] Standard errors are double-clustered by origin and destination states.

In our main specification, presented in Table 3 in the main body, the outcome variable is the value of trade. We measure this as the shipment value between the originating and destination states. All patterns documented in the paper persist if we measure trade volume as shipment weight in tons, or in ton-miles (the shipment weight multiplied by the mileage traveled by the shipment). The estimated coefficients are presented in Panel A of Appendix Table A4. Columns 1-4 are the same as in Panel A of Table 3 in the main body, and are based on the log value of shipment. In columns 5-6, we provide estimates based on the log ton of the shipment. Social connectedness also appears to be a significant explanatory variable for state trade flows when we measure trade by shipment weight in tons.

Binned Scatter Analysis. Appendix Figure A26 shows binned scatter plots that visualize the relationships between trade flows and geographic distance (Panel A), and between trade flows and social connectedness, conditional on geographic distance (Panel B). These plots complement the regression analyses in the main body of the paper and in Appendix Table A4, and highlight that the relationship between social connectedness and trade is essentially log-linear.

Analysis by Commodity Sector. To understand why trade volume is correlated with the strength of social connectedness, we further examine the variation of friendship elasticities of trade across major commodity sectors, and analyze how these elasticities vary with the labor and skill intensities of these sectors.[4] Specifically, we estimate equation A4 for each of the 23 major sectors in the CFS data. Panel A of Appendix Figure A27 shows a scatter plot of the friendship elasticities of trade with the labor intensities of each sector, measured as the share of labor compensation in the total cost of labor and

[2]For trade flows within a state, we follow Anderson and van Wincoop (2004) and measure the "geographic distance" between origin and destination as 0.25 times the distance to the nearest state.

[3]GDP per capita is obtained from the Bureau of Economic Analysis; the unemployment rates are obtained from the Bureau of Labor Statistics; union shares and population density are from Chodorow-Reich et al. (2012); and the sectoral composition is defined as the share of employees on non-farm payrolls in each major sector, obtained from the Bureau of Labor Statistics. The major sectors include construction, manufacturing, transportation, utilities, financial industry, professional services, education, health care, leisure, and government.

[4]We define commodity sectors based on the STCG trade sector categories in the CFS. We obtain the labor compositions of each sector using data from the EU KLEMS data. To merge these data sets, we manually map the SCTG codes in the CFS to SIC sector codes used in the EU KLEMS data. For example, we group all food products, such as the trade of cereal grains, milled grains, prepared food, and other food products, into one category. The final categories include food and beverages, agriculture, fishing, tobacco, mining of metal ores, mining of coal, extraction of gas and petroleum, quarrying, other non-metallic metals, chemical products (excluding pharmaceutical products), pharmaceutical products, wood products, pulp and paper products, printing and publishing production, metal goods, electrical equipment, machinery, transportation equipment, medical equipment, textiles, other miscellaneous manufacturing, rubber and plastics, and recycling products.

capital. There is no discernible correlation between the elasticities and labor intensities. Panel B of Appendix Figure A27 shows a scatter plot of the friendship elasticities with the share of high-skilled workers in each sector. The magnitude of the elasticity of trade flows with respect to friendship links rises with the share of high-skilled workers in the sector (the slope of the linear regression is 0.40, with a standard error of 0.21). These patterns are not driven by differences in the gender compositions of each sector. We find a positive relationship between the friendship elasticities and the share of both high-skilled male and female workers.

Sectors that have a larger share of high-skilled workers include those producing chemicals, pharmaceuticals, and medical equipment. One common characteristic of these sectors is that they produce products that are typically customized. In contrast, the sectors with a lower share of high-skilled workers often produce more standardized products such as wood, rubber, and plastics. One hypothesis consistent with these patterns is that informational asymmetries associated with the quality of the product can arise disproportionately with less standardized products. Social connectedness may help alleviate these information frictions, providing an explanation for the stronger positive relationship between trade and friendship links in these sectors. Investigating these and other channels through which trade patterns and friendship links are related is an exciting area for future research facilitated by the availability of the SCI data.

C.2 Social Connectedness and Patent Citations

Data Description. As we described in the main body of the paper, we start with all patents granted in the US in 2014. For each of these 2014 patents, we create an observation for every patent cited by the 2014 patent, so that the unit of observation is a patent-citation pair. For example, if a particular 2014 patent cites 10 other patents, this will generate 10 patent-citation pairs. We then construct a control observation for each of these patent-citation pairs. In particular, for each 2014 patent A that cites a previous patent B, we randomly select another 2014 patent C that is in the same technology class as patent A, but that does not cite patent B. We focus on patent classes with at least 1,000 patents issued in 2014, to ensure that there is a sufficient sample to randomly select the control patents.

Regression Analysis. Panel B of Table 3 in the main body explored the relationship between social connectedness and the probability of patent citations. These coefficients were estimated using equation A5:

$$C_{ij} = \beta_1 \log(d_{c(i)c(j)}) + \beta_3 \log(f_{c(i)c(j)}) + \beta_3 X_{c(i)c(j)} + \psi_{c(i)} + \psi_{c(j)} + \psi_{s(i)} + \psi_{s(j)} + \epsilon_{ij}. \tag{A5}$$

The unit of observation is a patent i-patent j pair. The counties where patents i and j were granted are denoted by $c(i)$ and $c(j)$, respectively. The technological class of patents i and j are denoted by $s(i)$ and $s(j)$, respectively. The dependent variable C_{ij} equals one if the issued patent i cites patent j, and zero otherwise. The variables $\log(d_{c(i)c(j)})$ and $\log(f_{c(i)c(j)})$ denote the log of geographic distance and log of the SCI between the counties of the issued and cited patents, respectively. The variable $X_{c(i)c(j)}$ denotes a vector of differences between the counties along the following dimensions: 2008 vote share of Obama, mean income, share of population without a high school degree, share of population that is white, share of population that is religious, and share of workforce employed in manufacturing. We also include fixed effects for the technology class of patents i and j (given by $\psi_{s(i)}$ and $\psi_{s(j)}$, respectively) and for the counties of of patents i and j (given by $\psi_{c(i)}$ and $\psi_{c(j)}$, respectively). We double cluster the standard errors by the technology classes of patents i and j.

Columns 1-3 and column 6 of Appendix Table A4, Panel B, are the same as columns 1-4 in Table 3, Panel B, in the main body of the paper. These columns show that the probability of citation falls as the distance between the counties of the issued and cited patents increases. Columns 4 and 5 show that these results are robust to variations in the controls included in the regression.

Binned Scatter Analysis. Appendix Figure A28, Panel A, shows binned scatter plots of the probability that the 2014 patent cites another patent against the log of the geographic distance between the counties of the issued and cited patents. We also control for fixed effects of the patent classes of the 2014 patent and the cited patent, and for county fixed effects. The average probability of citation is 0.5 by construction, since for each citing patent we included one non-citing control patent. Consistent with the existing literature, we find that the probability of a patent citation declines with geographic distance. In Panel B of Appendix Figure A28, we plot the probability of a patent citation against the log of the SCI, conditional on the same fixed effects as in Panel A, and also controlling flexibly for the log of the geographic distance between the counties. We find that the probability of a patent citation rises with the degree of social connectedness between the counties of the issued and cited patents, even after controlling for the geographic distance between these counties. The relationships is essentially log-linear. These results complement the regression analyses from the main paper.

C.3 Social Connectedness and Migration

Data Description. In the main body of the paper, we document that the social connections between two regions, as measured by the SCI, are strongly correlated with the extent of population flows between these regions. We measure migration using the SOI Tax Stats Migration Data provided by the IRS, which are based on year-to-year address changes reported on individual income tax returns. We focus on the migration of heads of households between 2013 and 2014, and calculate gross migration rates between each county-pair. One challenge with these data is that the IRS only reports flows for county-pairs with at least 20 movers; this corresponds to just over 25,000 county-pairs. The analysis in Panel C of Table 3 in the paper focuses on those county-pairs.

Regression Analysis. Panel C of Table 3 in the main body of the paper explored the relationship between migration and social connectedness. We analyze the relationship between social connectedness and population flows using the specification in equation A6. The dependent variable $\log(m_{ij})$ captures the log of total migration between counties i and j. The variable $\log(d_{ij})$ denotes the log of the geographic distance between counties i and j, and the variable $\log(f_{ij})$ denotes the log of the relative number of friendship links (i.e., the log of the SCI) between those counties. We also include fixed effects for each county, which allows us to control for the size of their populations and other county-level characteristics that might affect the degree of migration. Standard errors are double clustered at the levels of the counties within the county-pair.

$$\log(m_{ij}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \epsilon_{ij} \tag{A6}$$

Columns 1, 3, and 4 of Appendix Table A4, Panel C, are the same as those in Panel C of Table 3 in the main body of the paper. These columns show that individuals are much more likely to move to counties where they already have friends. In the other columns, we show that the results are robust to controlling more flexibly for the geographic distance between the counties, by including 50 dummies for equal-sized percentiles of the distance distribution.

Binned Scatter Analysis. Panel A of Appendix Figure A29 shows a binned scatter plot that documents the relationship between geographic distance and cross-county migration, controlling for county fixed effects. The relationship is not perfectly linear, with a somewhat more negative elasticity at shorter distances (this is the reason we included more flexible distance controls in regression A6, as described above). Panel B of Appendix Figure A29 shows a binned scatter plot that documents the relationship between the SCI and migration flows non-parametrically, controlling for county fixed effects and the geographic distance between county-pairs: the relationship is almost log-linear, suggesting a constant elasticity between friendship links and migration.

State-State Job Flows. In the main body, we study the effect of social connectedness on within-US migration using county-level data from the SOI Tax Stats Migration Data provided by the IRS. A second data set to analyze within-US migration comes from the US Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) database. The sample spans from Q2 2000 to Q2 2014. We examine the correlation of social connectedness with US state-state quarterly job flows using the publicly available LEHD data, which give the count of job transitions between states. The LEHD data is also disaggregated by firm characteristics (industry, age, and size of the origin and destination firms), and worker demographics (gender by age, gender by education, and race), which can be used to explore heterogeneity in the importance of social connectedness in facilitating labor flows.

First, we analyze the correlation between job flows and friendship links using the specification in equation A7 below. The dependent variable $\log(y_{ijt})$ captures the log number of gross job flows between states i and j in quarter t. The variable $\log(d_{ij})$ denotes the log of the geographic distance between states i and j, and the variable $\log(f_{ij})$ captures the log of the SCI between the states. We include time fixed effects, denoted by ψ_t, to control for common aggregate shocks affecting all states. We also include fixed effects for origin and destination states, denoted by ψ_i and ψ_j. Finally, we include a dummy variable if the two states are adjacent to each other, and a dummy variable to capture within-state flows. In some specifications, X_{ij} includes other non-time-varying controls for differences between states i and j, such as the differences in the relative levels of GDP per capita, unemployment rates, union share, population density, and sectoral composition.

$$\log(y_{ijt}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \psi_t + \epsilon_{ijt} \tag{A7}$$

Appendix Table A5 reports the estimates from equation A7, focusing on the within-quarter job flows.[5] Column 1 does not control for the SCI. The estimated elasticity of job flows with respect to geographic

[5] We obtain similar results using the other measures of job flows. These other measures include job flows when the job transition did not occur within the same quarter. We also observe similar results when we analyze job flows between "stable" jobs, which the Census defines as the jobs that are held on the first and the last day of the quarter.

distance is very close to -1, resembling the migration elasticity estimates from the county-county migration data. Panel A of Appendix Figure A30 shows a binned scatter plot that documents the nearly linear relationship non-parametrically, controlling for state fixed effects.

In column 2 of Appendix Table A5, we do not control for geographic distance, but instead include the log of the SCI. The estimated elasticity of state-level job flows to friendship links is about 1.15, which is again very close to the elasticity of county-level migration to friendship links. Panel B of Appendix Figure A30 shows a binned scatter plot that documents the strong and nearly log-linear relationship between the SCI and labor flows non-parametrically, controlling for state fixed effects and geographic distance. If we control for both geographic distance and social connectedness in column 3 of Appendix Table A5, we observe that the SCI-elasticity of labor flows declines only slightly, while the distance-elasticity of labor flows drops by 80%. This pattern is highly consistent with the results from the county-level migration analysis. In column 4 of Appendix Table A5, we further control for other differences between the two states (described above). This has a negligible effect on the R^2, and no effect on the estimates of distance-elasticity and SCI-elasticity of labor flows. Finally, in column 5 we include fixed effects for origin and destination state interacted with the calendar quarter. This controls, for example, for time-variation in the economic conditions in the states. While the R^2 increases somewhat, our estimates of interest remain unaffected.

In addition to considering the average elasticity of labor flows to social connectedness, we also analyzed whether there was any heterogeneity in this elasticity along characteristics of the new firm that employs the person switching states. Interestingly, we find no differences in the labor flow elasticities to social connectedness along the age or the size of the destination firm.

D International Dimension of Social Connectedness of US Counties

In this Appendix, we explore a number of additional dimensions of social connectedness of US counties to foreign countries.

Explorations of International Social Connectedness. Appendix Figure A31 shows the share of friendship links of each county in the continental United States to various foreign countries. Panels A and B show linkages to **Canada** and **Mexico**, which are mostly concentrated on the land-borders with those states. Panel C shows the share of friendship links to **Norway**. As discussed in the main body of the paper, these are particularly strong in those Midwestern states that saw significant Nordic immigration. Similarly, Panel D shows the distribution of friendship links to **Germany**. Relative connections to Germany are particularly strong for those counties in the Midwest (particularly in Michigan, Wisconsin, and Minnesota) that saw major German immigration from these countries in the late 19th Century. The county with the strongest connection to Germany is Otero County, New Mexico, home to the German Air Force Flying Center at Holloman Air Force Base. Panel E shows connections to **Italy**. These are particularly strong in the Northeastern United States, a region with strong Italian immigration. The SCI also allows us to identify present-day links to relatively small countries with more limited migration to the United States. There are roughly 100,000 Americans of Somali descent, mostly living in Minnesota. Panel F shows that this region is also strongly connected to **Somalia** in the SCI data. The shaded region in Colorado surrounds the town of Fort Morgan, where roughly one in ten of the 12,000 residents is Somali (see Denver Post, 2011, for more information). These patterns suggest a strong link between present-day social connectedness and past migration, which we explore in more detail in the next section. Panel G shows the distribution of friendship links to **South Africa**. There is a region of high connectivity in Montana and North Dakota, likely related to the significant movement of South African farmhands to this region (see Grand Forks Herald, 2014; Great Falls Tribune, 2016, for more information). Panel H shows the share of friendship links to **Cape Verde**, which are particularly notable in New England. This is consistent with the long history of Cape Verdean Americans settling in Massachusetts. Indeed, of the roughly 115,000 Americans who report Cape Verdean ancestry, about 75,000 are in Massachusetts alone (see U.S. Census Bureau, 2016, for demographic data from the Census and the American Community Survey). Panel I shows the same measure for the island nation of **Kiribati**, revealing that the region with the highest share of friendship links to Kiribati is in and surrounding Utah. Kiribati, along with many other Pacific island countries, has a large Mormon population (see Mormon Newsroom, 2016, for more information). Panels J and K show the share of friendship links to **Cambodia** and **Laos**, respectively, by county. Both show strong ties to the West Coast, where most Americans of Cambodian and Laotian descent live; to Massachusetts, which accepted a large number of refugees from both countries; and to the Washington, DC, area. The high share of links to Laos in Minnesota and Wisconsin likely reflects Laotian refugee resettlement in the state, and the pocket of connections in Arkansas may reflect Arkansas' status as an entry point for Indochinese refugees (see Wikipedia, 2016; Lao Assistance Center of Minnesota, 2016; Encyclopedia of Arkansas History and Culture, 2015, for more information). Cambodia shows a higher share of friendship links in Texas, where the 2010 Census counted 14,000 people of Cambodian descent. Dallas and Houston are home to the largest Cambodian communities not in Massachusetts, the Washington, DC, area, or on the West Coast (see Khmer Salem Blog, 2013, for more information). Panel L shows connections to **Ethiopia**, including prominent population centers in the Washington, DC, area and Minnesota. The Washington, DC, area is home to the largest concentration of people of Ethiopian descent outside of Africa, and Minnesota is, as also illustrated in Panels D and E, a major destination for refugees (see WAMU, 2016; Twin Cities World Refugee Day, 2016, for more information).

Appendix Figure A32 shows county-level binned scatter plots of the relationship between the number of respondents with ancestry from a given country and today's social connectedness with that country. These plots are shown for four countries for which immigration peaked at different times: Ireland, Italy, Greece, and the Philippines. We control flexibly for the log of the geographic distance between the county and the foreign country's capital. There is a strong relationship between ancestry from a given country and the extent of present-day social connectedness, even for those countries with immigration waves that peaked more than 100 years ago.

Further Details on Table 4. In columns 1-3 of Table 4 in the main paper, we formally analyze the extent to which past migration from a particular country is correlated with the strength of today's social connectedness with that country. We use two measures of past migration: the number of residents in each county who were born in a specific foreign country, and the number of residents who claim their primary ancestry as being from a given country. We obtain individual-level information on these two variables from the 2014 5-year ACS, and aggregate these measures to the county level. County identifiers are available for counties with a sufficient number of residents to guarantee anonymity, leaving us with 473 counties. Throughout the analysis, we restrict our attention to foreign countries from which at least 800 respondents across the US claim ancestry.[6] In the data, most foreign countries are coded individually for ancestry; for foreign birthplace, some of these are pooled together into broader categories, such as South America.

Columns 1-3 of Table 4 in the main body present estimates from the following regression equation to analyze the effect of past migration on current social connectedness:

$$log(f_{ic}) = \beta_1 log(anc_{ic}) + \beta_2 log(d_{ic}) + \psi_c + \psi_i + \epsilon_{ic}, \tag{A8}$$

where f_{ic} is the SCI between county c and foreign country i, anc_{ic} is the number of individuals in county c who stated their ancestry as being of country i, and d_{ic} is the geographic distance between county c and country i's capital city. We also include fixed effects for each county and foreign country. There are many counties which do not have a resident with a given country's ancestry. To include county-country pairs with zero ancestry links in our analysis, columns 4-6 of Panel A, Table A6 also estimate the above equation with $log(f_{ic} + 1)$ as the dependent variable; in those specifications, the key explanatory variable is $log(anc_{ic} + 1)$. The overall message remains very similar.

We also analyze whether the elasticity of friendship links to ancestry varies by when immigration to the US peaked. To measure this, we determine the Census year in which the number of recent immigrants from each country was the highest. We then split countries into four groups: peak immigration in the 1890 Census, in the 1910 to 1930 Census, in the 1960 to 1990 Census, and in the 2000 Census. Figure A33 shows estimated coefficients for β_1 separately for counties by their immigration peaks. The elasticity of social connectedness to ancestry declines in the time since the peak of the immigration wave, but remains strong even for countries with little recent migration.

Social Connections and International Trade. We next analyze the extent to which higher social connectedness with foreign countries is associated with more trade with these countries. This mirrors the analyses in the main body, which estimated the role of social connectedness in facilitating within-US trade. This analysis contributes to a literature that has documented the significant extent to which past international migration can help facilitate present-day economic interactions, such as trade and foreign direct investment, with the origin countries of these migrants (e.g., Gould, 1994; Rauch and Trindade, 2002; Felbermayr, Grossmann and Kohler, 2012; Parsons and Vézina, 2014; Burchardi, Chaney and Hassan, 2016). While this literature has made progress in establishing causal relationships, often using quasi-exogenous variation in the destination of migrants, the precise channel underlying any observed relationships is less clear. For example, one important reason why trade might increase with past migration is because such migration can lead to higher present-day social connectedness with the origin country of the migrants, which can help alleviate informational frictions; however, other possible channels could be, among others, common tastes (Gould, 1994) or higher trust between culturally more similar individuals (Guiso, Sapienza and Zingales, 2009). The literature has struggled to separate these mechanisms, in part because of the challenges of measuring present-day social connectedness between US regions and foreign countries. The SCI provides such a measure.

Since no data on trade at the county level are publicly available, we focus on measuring inter-

[6] Our results are robust to varying this cut-off or including all countries. We also re-code ancestry and foreign birthplaces to match today's political boundaries, such as coding "Persian" ancestry to "Iran."

national trade at the state level. Data on international trade by state and foreign trading partner are obtained from the International Trade Administration. We focus on the value of total imports and exports in 2015, measured in US dollars. Specifically, we estimate the relationship between today's social connectedness and imports and exports with the following regression:

$$log(t_{is}) = \beta_1 log(f_{is}) + \beta_2 log(d_{is}) + \psi_s + \psi_i + \epsilon_{is}, \tag{A9}$$

where t_{is} is trade between state s and country i, f_{is} is the SCI, and d_{is} is the distance between capital cities. We also include fixed effects for each state and foreign country. There are many country-state pairs without any trade. To include these in the regression, we also estimate the above equation with $log(t_{is} + 1)$ as the dependent variable, replacing $log(f_{is})$ by $log(f_{is} + 1)$.

Panel B of Table A6 shows that including the SCI as an explanatory variable reduces the estimated relationship between distance and international trade substantially, by about one third for imports and a quarter for exports. Social connectedness itself strongly correlates with the volume of international trade. A state with 10% higher connectivity to a given foreign country on average imports 3.1% more from this country and exports 3.4% more to this country. Including state-country pairs with no imports or exports in 2015 by estimating equation A9 with $log(t_{is} + 1)$ instead of $log(t_{is})$ as the dependent variable increases the estimated effects of social connectivity. In this specification, 10% higher connectivity is associated with 4.7% higher imports and 6% higher exports.

References

Amauta. 2010. "American Indian and Alaskan Native Population." http://www.amauta.info/maps/aipopulation.jpg, accessed: 2016-12-21.

Anderson, James E, and Eric van Wincoop. 2004. "Trade costs." *Journal of Economic Literature*, 42(3).

Burchardi, Konrad, Thomas Chaney, and Tarek Hassan. 2016. "Migrants, Ancestors, and Investments."

Chetty, Raj, and Nathaniel Hendren. 2015. "The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates." *Unpublished Manuscript*.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The association between income and life expectancy in the United States, 2001-2014." *JAMA*.

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): 1553–1623.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston. 2012. "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy*, 4(3): 118–45.

Crew, Spencer R. 1987. "The Great Migration of Afro-Americans, 1915-40." *Monthly Labor Review*, 110(3): 34–36.

Denver Post. 2011. "Somali Muslims finding a place in Fort Morgan." http://www.denverpost.com/2011/08/16/somali-muslims-finding-a-place-in-fort-morgan/, accessed: 2016-12-21.

Dial, Wylene P. 1969. "The dialect of the Appalachian people." *West Virginia History*, 30(2): 463–471.

Encyclopedia Britannica. 2012. "Map of the Blue Ridge Mountains." http://media.web.britannica.com/eb-media/55/89855-004-33F625A4.gif, accessed: 2016-12-21.

Encyclopedia of Arkansas History and Culture. 2015. "Indochinese resettlement program." http://www.encyclopediaofarkansas.net/encyclopedia/entry-detail.aspx?entryID=5562, accessed: 2016-12-21.

Felbermayr, Gabriel, Volker Grossmann, and Wilhelm Kohler. 2012. "Migration, International Trade and Capital Formation: Cause or Effect?"

Gould, David M. 1994. "Immigrant links to the home country: empirical implications for US bilateral trade flows." *The Review of Economics and Statistics*, 302–316.

Grand Forks Herald. 2014. "South Africans learn new farming techniques, enjoy life in Lankin." http://www.greatfallstribune.com/story/money/2016/02/12/montana-farmers-turn-south-africa-fill-labor-gap/80295960/, accessed: 2016-12-21.

Great Falls Tribune. 2016. "Montana farmers turn to South Africa to fill labor gap." http://www.grandforksherald.com/content/south-africans-learn-new-farming-techniques-enjoy-life-lankin, accessed: 2016-12-21.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics*, 124(3): 1095–1131.

Holt-Lunstad, Julianne, Timothy B Smith, and J Bradley Layton. 2010. "Social relationships and mortality risk: a meta-analytic review." *PLoS Med*, 7(7): e1000316.

House, James S, Karl R Landis, Debra Umberson, et al. 1988. "Social relationships and health." *Science*, 241(4865): 540–545.

Infogroup. 2009. "Data dictionary: Religion 2009." https://www.socialexplorer.com/data/Religion_InfoUSA09/metadata/?ds=TR, accessed: 2016-12-21.

Khmer Salem Blog. 2013. "What U.S. city has the most Khmer?" http://khmersalem.blogspot.com/2013/06/what-us-city-has-most-khmer.html, accessed: 2016-12-21.

Lambiotte, Renaud, Vincent D Blondel, Cristobald De Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. 2008. "Geographical dispersal of mobile communication networks." *Physica A: Statistical Mechanics and its Applications*, 387(21): 5317–5325.

Lao Assistance Center of Minnesota. 2016. "Homepage." http://laocenter.org/, accessed: 2016-12-21.

Liben-Nowell, David, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2005. "Geographic routing in social networks." *Proceedings of the National Academy of Sciences of the United States of America*, 102(33): 11623–11628.

Marathon Petroleum. 2016. "Illinois Refining Division." http://www.marathonpetroleum.com/Operations/Refining_and_Marketing/Refining/Illinois_Refining_Division/, accessed: 2016-12-21.

Mormon Newsroom. 2016. "Worldwide Statistics: Kiribati." http://www.mormonnewsroom.org/facts-and-statistics/country/kiribati, accessed: 2016-12-21.

National Park Service. 2003. "Indian reservations in the continental United States." https://www.nps.gov/nagpra/DOCUMENTS/RESERV.PDF, accessed: 2016-12-21.

New York Times. 2008. "The dozens of languages of the Caucasus say much about the Georgia conflict." http://www.nytimes.com/2008/08/24/world/europe/24iht-caucasus.4.15591770.html, accessed: 2016-12-21.

NPR. 2015. "Some anxiety, but no slowdown for North Dakota oil boom town." http://www.npr.org/2015/03/20/393639392/some-anxiety-but-no-slowdown-for-north-dakota-oil-boom-town, accessed: 2016-12-21.

Parsons, Christopher, and Pierre-Louis Vézina. 2014. "Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment."

Rauch, James E, and Vitor Trindade. 2002. "Ethnic Chinese networks in international trade." *Review of Economics and Statistics*, 84(1): 116–130.

Rupasingha, Anil, Stephan J. Goetz, and David Freshwater. 2006. "The production of social capital in US counties." *The Journal of Socio-Economics*, 35(1): 83 – 101. Essays on Behavioral Economics.

Scellato, Salvatore, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. 2011. "Socio-Spatial Properties of Online Location-Based Social Networks." *ICWSM*, 11: 329–336.

The Guardian. 2009. "2008 presidential election results by state and county." *https://www.theguardian.com/news/datablog/2009/mar/02/us-elections-2008*, accessed: 2016-12-21.

Tolbert, Charles M, and Molly Sizer. 1996. "US commuting zones and labor market areas."

Tolnay, Stewart E. 2003. "The African American "Great Migration" and beyond." *Annual Review of Sociology*, 29: 209–232.

Twin Cities World Refugee Day. 2016. "Refugees in Minnesota." *http://tcworldrefugeeday.org/aboutrefugees-2/*, accessed: 2016-12-21.

U.S. Census Bureau. 2016. "American Fact Finder." *https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t*, accessed: 2016-12-21.

WAMU. 2016. "Why is there such a large Ethiopian population in the Washington region?" *http://wamu.org/story/16/04/21/how_did_the_dc_region_become_home_to_the_largest_population_of_ethiopians_in_the_us/*, accessed: 2016-12-21.

Wikimedia. 2010. "Great Appalachian Valley map." *https://commons.wikimedia.org/wiki/File:Greatvalley-map.png*, accessed: 2016-12-21.

Wikipedia. 2016. "Hmong in Wisconsin." *https://en.wikipedia.org/wiki/Hmong_in_Wisconsin*, accessed: 2016-12-21.

Appendix Tables and Figures

Table A1: Distance and Friendship Links: Across-County Summary Statistics

	Share of Friends Living Within:				Share of U.S. Population Living Within:			
	50 Miles	100 Miles	200 Miles	500 Miles	50 Miles	100 Miles	200 Miles	500 Miles
Mean	55.4%	62.8%	70.3%	79.7%	1.3%	2.8%	6.6%	22.3%
P5	38.1%	46.0%	54.2%	64.2%	0.1%	0.3%	1.0%	5.5%
P10	42.5%	49.6%	57.1%	66.7%	0.1%	0.6%	2.1%	7.9%
P25	48.4%	55.9%	63.8%	74.6%	0.3%	1.1%	3.5%	13.9%
Median	55.4%	63.9%	71.6%	81.9%	0.7%	2.1%	5.8%	22.5%
P75	63.2%	70.9%	78.0%	86.2%	1.8%	3.5%	8.2%	30.7%
P90	67.4%	74.8%	81.2%	89.0%	3.2%	6.2%	15.0%	37.1%
P95	70.3%	76.9%	83.2%	91.0%	5.4%	9.2%	15.6%	39.7%

Note: Table shows across-county summary statistics for the share of friends of the county's population living within a certain distance of that county, and the share of the US population living within a certain distance. Counties are weighted by their population.

Table A2: Determinants of Social Connectedness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Pop_1 x Pop_2)	1.075***	1.133***					
	(0.021)	(0.019)					
Log(Distance in Miles)		-1.067***	-1.483***	-1.287***	-1.160***	-1.988***	-1.214***
		(0.064)	(0.065)	(0.061)	(0.059)	(0.043)	(0.055)
Same State				1.496***	1.271***	1.216***	1.496***
				(0.087)	(0.083)	(0.044)	(0.085)
Δ Income (k$)							-0.006***
							(0.001)
Δ Share Pop White (%)							-0.012***
							(0.001)
Δ Share Pop No High School (%)							-0.012***
							(0.002)
Δ 2008 Obama Vote Share (%)							-0.006***
							(0.001)
Δ Share Pop Religious (%)							-0.002***
							(0.001)
County Fixed Effects	N	N	Y	Y	Y	Y	Y
Sample					>200 miles	<200 miles	
Number of observations	2,961,970	2,961,970	2,961,968	2,961,968	2,775,244	186,669	2,961,968
R2	0.682	0.813	0.907	0.916	0.916	0.941	0.922

Note: Table shows results from regression A2. The unit of observation is a county-pair, the dependent variable is the log of the SCI. Standard errors are double clustered at the level of the states of the two counties, and are given in parentheses. Details on the data construction for demographic differences can be found in the Online Appendix. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Table A3: Concentration of Networks and Socioeconomic Outcomes

	Share of Friends Within 100 Miles			Share of Friends Among Nearest 50 Million People		
	(1)	(2)	(3)	(4)	(5)	(6)
Average Income (k$)	-0.122***	-0.180***	-0.279***	-0.209***	-0.220***	-0.203***
	(0.017)	(0.015)	(0.019)	(0.010)	(0.009)	(0.012)
Causal Social Mobility	1.720**	4.216***	5.736***	2.917***	3.230***	3.631***
	(0.678)	(0.638)	(0.668)	(0.373)	(0.382)	(0.419)
Social Capital	-0.513*	-1.555***	-0.572*	0.124	0.279	0.139
	(0.287)	(0.295)	(0.342)	(0.158)	(0.176)	(0.214)
Life Expectancy at Q1 Income	-1.852***	-0.845***	-0.396**	-0.717***	-0.261**	-0.137
(Conditional on Race)	(0.181)	(0.170)	(0.197)	(0.100)	(0.102)	(0.123)
State Fixed Effects	N	Y	N	N	Y	N
Commuting Zone Fixed Effects	N	N	Y	N	N	Y
N	1,546	1,545	1,375	1,546	1,545	1,375
R-Squared	0.127	0.518	0.752	0.306	0.542	0.764

Note: Table shows results from a regression of county-level measures of the concentration of social networks on measure of county-level socioeconomic outcomes. In columns 1 to 3, we analyze the correlation with the share of friends living within 100 miles, in columns 4 to 6 the share of friends within the closest 50 million people. In columns 2 and 5 we also control for state fixed effects, and in columns 3 and 6 we also control for commuting zone fixed effects.

Table A4: Social Connectedness and Across-Region Economic Interactions

	Panel A - Dependent Variable: State-Level Trade Flows					
	Log(Value)				Log(Tons)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Distance)	-1.057***		-0.531***	-0.533***	-1.081***	-1.044***
	(0.071)		(0.084)	(0.085)	(0.097)	(0.101)
Log(SCI)		0.999***	0.643***	0.637***	0.721***	0.768***
		(0.051)	(0.071)	(0.060)	(0.089)	(0.102)
State Fixed Effects	Y	Y	Y	Y	Y	Y
Other State Differences	N	N	N	Y	N	Y
N	2,219	2,220	2,219	2,219	1,935	1,935
R-Squared	0.912	0.918	0.926	0.930	0.893	0.895

	Panel B - Dependent Variable: Patent Citation					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Distance)	-0.048***		-0.011**	-0.011**	-0.018**	-0.021**
	(0.002)		(0.005)	(0.005)	(0.008)	(0.009)
Log(SCI)		0.063***	0.049***	0.044***	0.060***	0.066***
		(0.003)	(0.006)	(0.006)	(0.010)	(0.012)
Technological Category Fixed Effects	Y	Y	Y	Y	Y	Y
County Fixed Effects	Y	Y	Y	Y	Y	Y
Other County Differences	N	N	N	Y	Y	Y
Cited Patent Fixed Effects	N	N	N	N	Y	Y
Issued (2014) Patent Fixed Effect	N	N	N	N	N	Y
N	2,171,754	2,171,754	2,171,754	2,168,790	2,168,370	2,168,285
R-Squared	0.056	0.059	0.059	0.060	0.085	0.101

	Panel C - Dependent Variable: County-Level Migration					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Distance)	-0.973***	Flexible		0.023	Flexible	Flexible
	(0.048)			(0.021)		
Log(SCI)			1.134***	1.148***	1.123***	1.134***
			(0.019)	(0.024)	(0.026)	(0.025)
County Fixed Effects	Y	Y	Y	Y	Y	Y
Other County Differences	N	N	N	N	N	Y
N	25,305	25,305	25,305	25,305	25,305	25,287
R-Squared	0.610	0.618	0.893	0.893	0.898	0.899

Note: Panel A shows results from regression A4. The unit of observation is a state-pair. The dependent variable in columns 1 through 4 is the log of the value of commodity flows between the states, and the log of the weight of commodity flows in column 5. All specifications include fixed effects for origin and destination state, as well as dummies for neighboring states and own-state flows (not shown). Columns 4 and 5 also control for differences between the states along the following dimensions: GDP per capita, unemployment rates, sectoral composition, union share, and population density. The standard errors are double-clustered by destination and origin states. Panel B shows results from regression A5. The unit of observation is a county-pair. The columns vary in the controls included in the specification. Standard errors are depicted in parentheses and are clustered by technology classes of the patent-citation pair. Panel C shows results from regression A6. The unit of observation is a county-pair. The dependent variable is the log of the gross migration of heads of households between the counties. All specifications include county fixed effects. Column 6 also controls for differences between the counties along the following dimensions: 2008 vote share of Obama, mean income, share of population without a high school degree, share of population that is white, share of population that is religious, and share of workforce employed in manufacturing. Standard errors are double clustered at the level of the states of the two counties, and are given in parentheses. See text for more details. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Table A5: Labor Flows and Social Connectedness

	(1)	(2)	(3)	(4)	(5)
Log(Distance)	-1.001***		-0.221***	-0.209***	-0.208***
	(0.062)		(0.022)	(0.022)	(0.022)
Log(SCI)		1.152***	1.007***	1.000***	1.001***
		(0.025)	(0.027)	(0.028)	(0.030)
Other State Differences	N	N	N	Y	Y
State and Time Fixed Effects	Y	Y	Y	Y	N
State x Time Fixed Effects	N	N	N	N	Y
N	120,374	120,431	120,374	120,374	120,374
R-Squared	0.922	0.957	0.959	0.959	0.971

Note: Table shows results from regression A7. The unit of observation is a state-pair-quarter. The dependent variable is the log of the job flows between the states. All specifications include fixed effects for origin and destination states, column 6 includes fixed effects for origin and destination state interacted with the quarter. In addition, all specifications include dummies for neighboring states and own-state flows (not shown). Columns 5 and 6 also control for differences between the states along the following dimensions: GDP, unemployment rates, sectoral composition, union share, and population density. The standard errors are double-clustered based on the destination and origin state. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

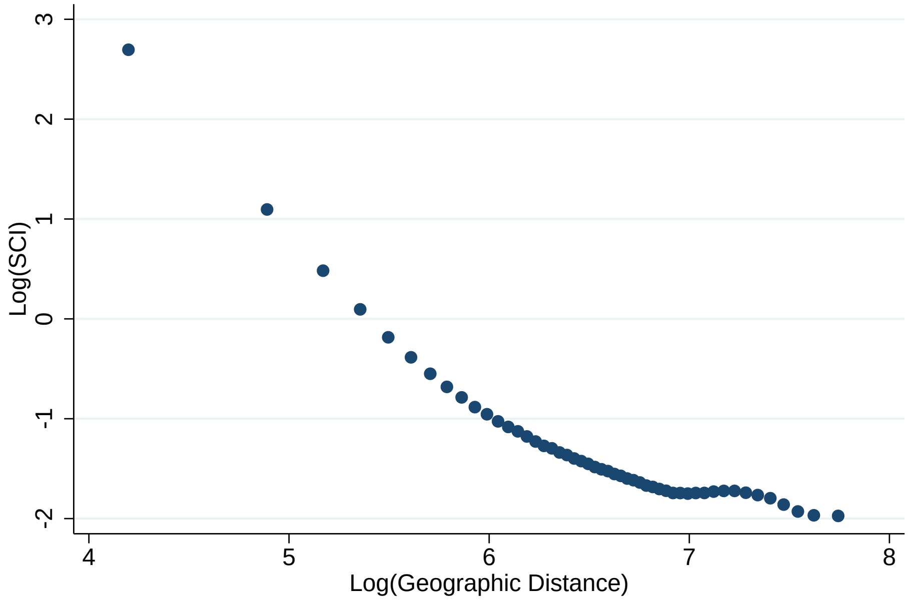
Table A6: The International Dimension of Social Connectedness

	Panel A: Ancestry and Social Connectedness					
	Log(SCI)			Log(SCI+1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Distance)	-1.159***	-0.690***	-0.493***	-1.113***	-0.653***	-0.471***
	(0.258)	(0.162)	(0.174)	(0.197)	(0.125)	(0.163)
Log(Ancestry in Foreign Country)		0.341***				
		(0.022)				
Log(Born in Foreign Country)			0.367***			
			(0.033)			
Log(Ancestry in Foreign Country+1)					0.352***	
					(0.022)	
Log(Born in Foreign County+1)						0.368***
						(0.029)
Fixed Effects	Y	Y	Y	Y	Y	Y
N	33,146	33,146	16,527	49,665	49,665	24,596
R-Squared	0.908	0.936	0.943	0.905	0.934	0.951
Number of Countries	105	105	52	105	105	52

	Panel B: Social Connectedness and International Trade					
	Log(Imports)		Log(Imports+1)	Log(Exports)		Log(Exports+1)
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Distance)	-1.535***	-1.064***	-1.627***	-2.038***	-1.506***	-2.092***
	(0.376)	(0.321)	(0.378)	(0.291)	(0.268)	(0.391)
Log(SCI)		0.313***	0.470***		0.338***	0.597***
		(0.075)	(0.103)		(0.053)	(0.139)
Fixed Effects	Y	Y	Y	Y	Y	Y
N	7,413	7,413	11,014	9,070	9,070	11,015
R-Squared	0.789	0.790	0.770	0.835	0.838	0.770
Number of Countries	212	212	216	215	215	216

Note: Panel A shows results from regression A8. The unit of observation is a US county-foreign country pair. Each specification also includes fixed effects for the US state and the foreign country. Standard errors are clustered at both the county and foreign country level. Panel B shows results from regression A9 for 2015 levels of total imports (columns 1 to 3) and 2015 levels of exports (column 4 - 6), both in current US Dollars. The unit of observation is a US state-foreign country pair. Each column also includes fixed effects for the US state and for the foreign country. Standard errors are clustered at both the state and foreign country level. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Figure A1: Geographic Distance and County-Level Social Connectedness

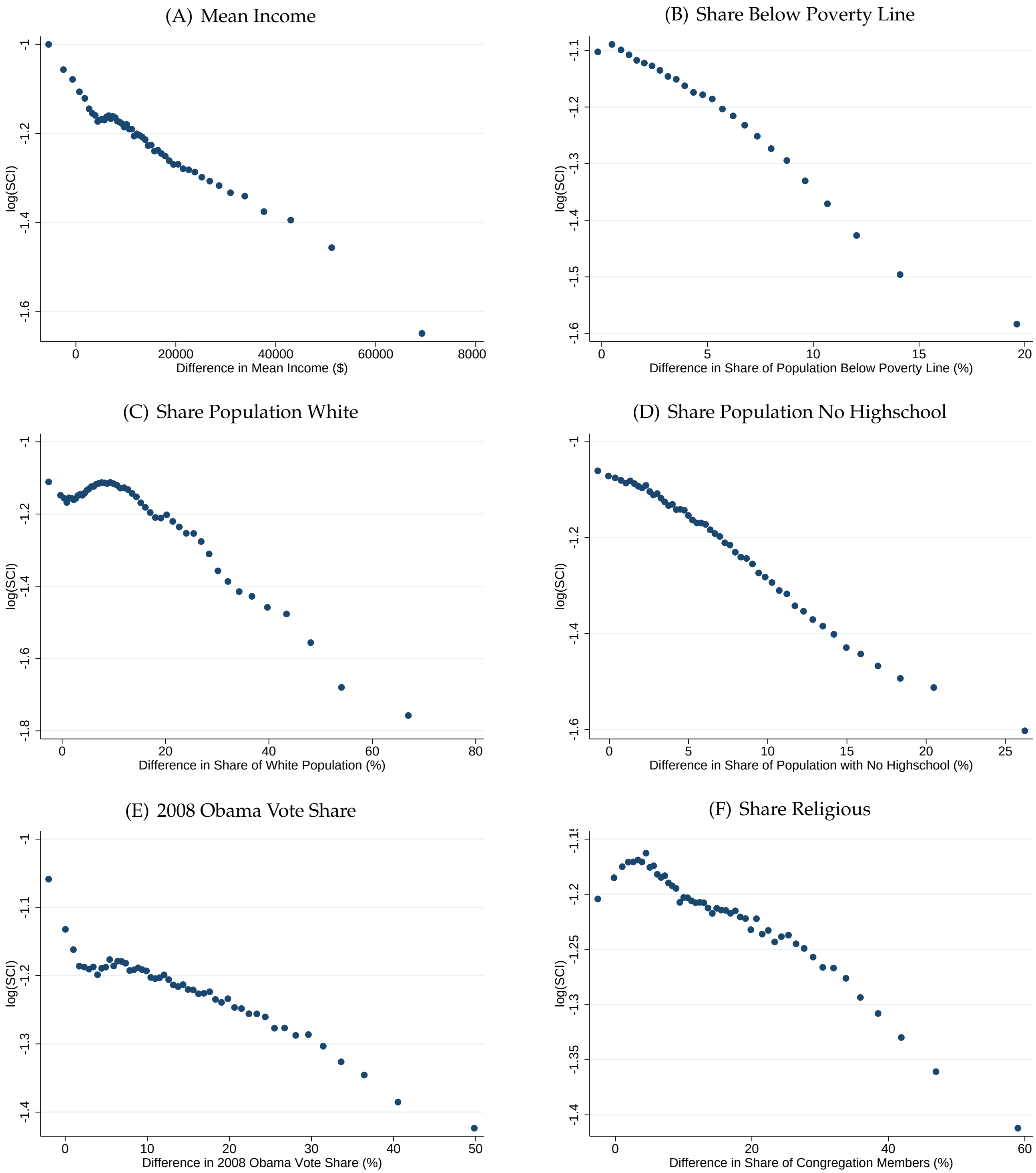


Note: Figure shows a conditional binned scatter plot with county-pairs as the unit of observation. On the horizontal axis is the log of the distance between the two counties, measured in miles, and on the vertical axis is the log of the SCI. We control flexibly for the log of the product of the populations in the two counties, by including 50 dummy variables for equal-sized quantiles of the distribution. We drop all county-pairs where either county has a population of fewer than 10,000 people.

Figure A2: State-State Adjacency Matrix of Friend Links

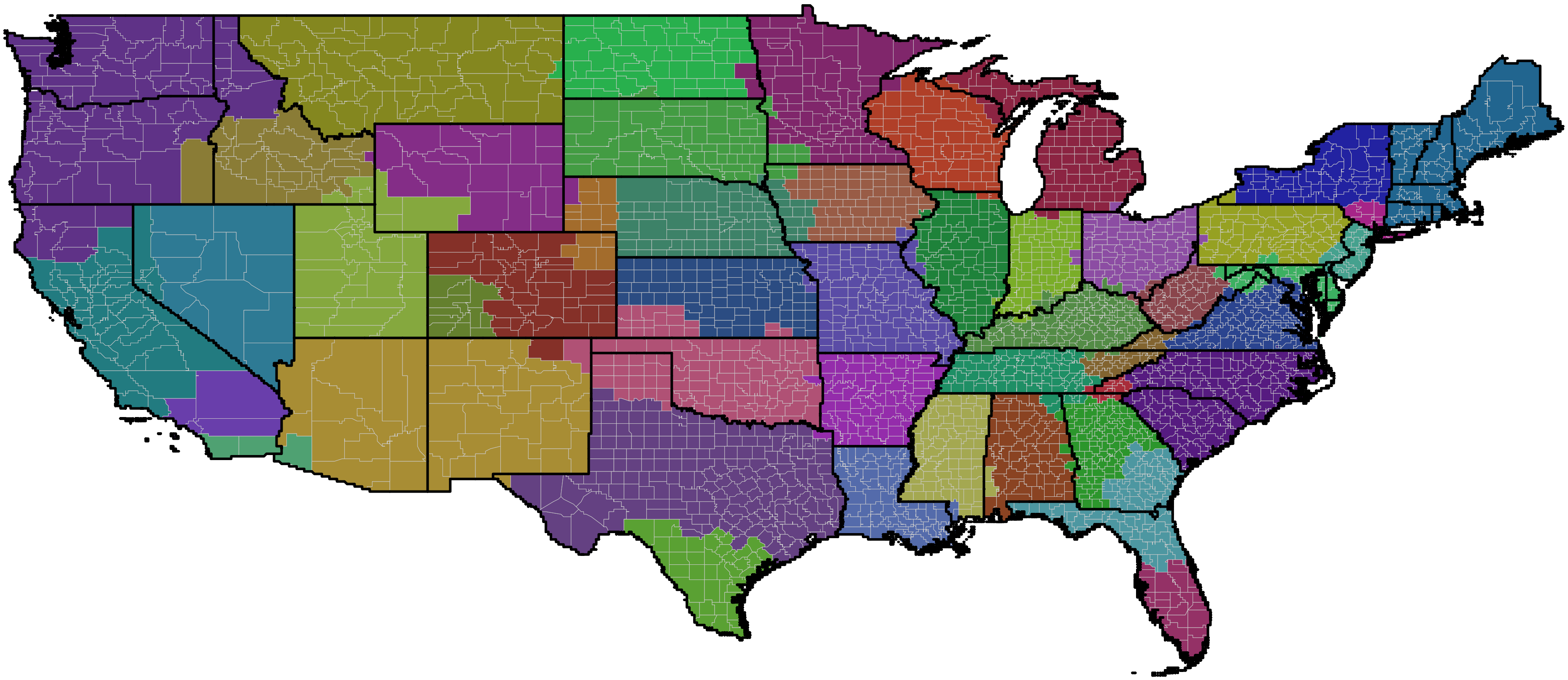
Note: Figure shows an adjacency matrix of the probability of social connections, constructed as in equation A1, and scaled as percentiles of connection strength. States are grouped by their Census Bureau Divisions (1 - New England; 2 - Middle Atlantic; 3 - East North Central; 4 - West North Central; 5 - South Atlantic; 6 - East South Central; 7 - West South Central; 8 - Mountain; 9 - Pacific).

Figure A3: The Geographic Spread of Friendship Networks and County-Level Outcomes



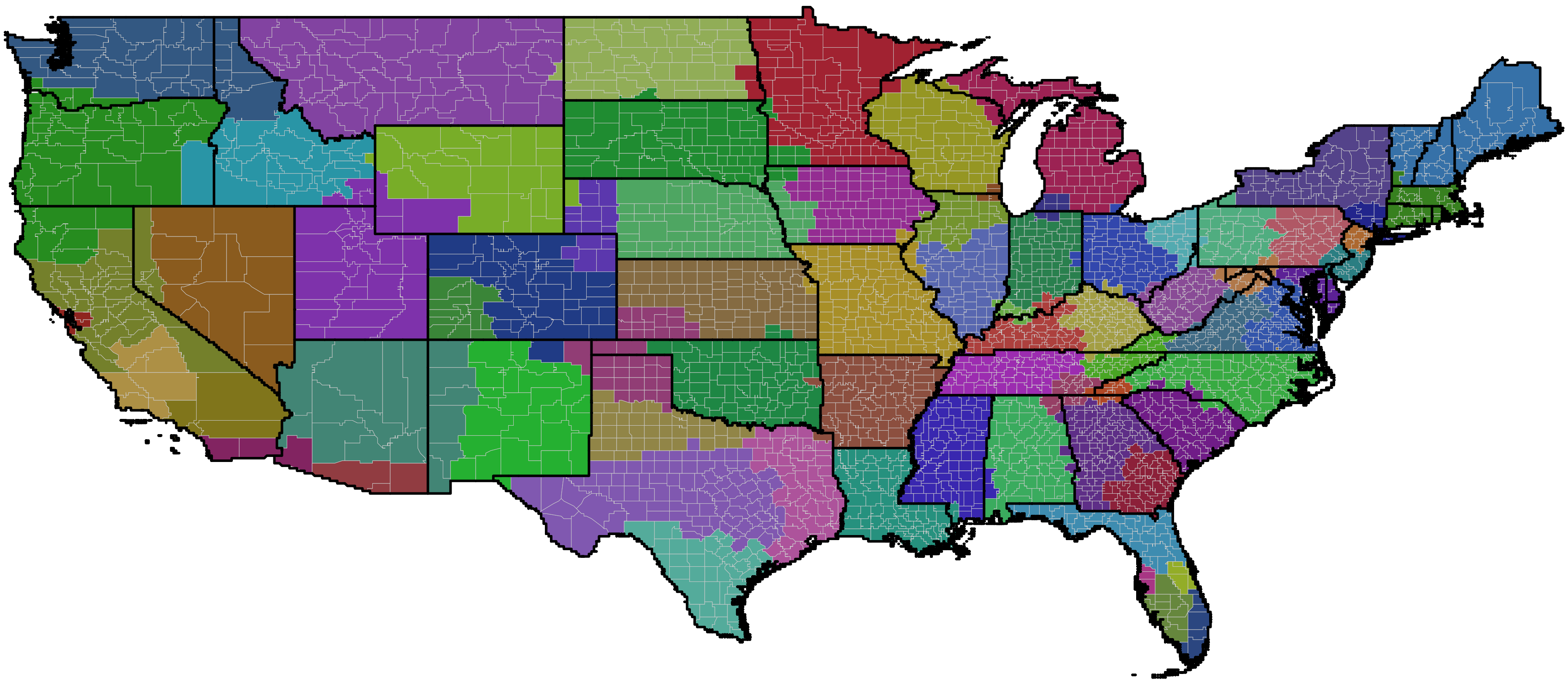
Note: Figure shows binned scatter plots, with county-pairs as the unit of observation. On the horizontal axis is the difference across the county-pairs for a number of county-level measures: mean income (Panel A), share of population below poverty line (Panel B), share of population that is white (Panel C), share of population with no high school degree (Panel D), the 2008 Obama vote share (Panel E), and the share of population that belongs to a major religious tradition's congregation (Panel F). On the vertical axis is the log of the number of friendship links between these counties, i.e., the log of the SCI. Each scatter plot controls flexibly for the log of the product of the counties' populations, and the log of the geographic distances between each county-pair.

Figure A4: Connected Communities within the United States - 50 Distinct Units



Note: Figure shows US counties grouped together when we use hierarchical agglomerative linkage clustering to create 50 distinct groups of counties. The algorithm assigns both Hawaii and Alaska, not pictured, to two distinct clusters including only the respective state.

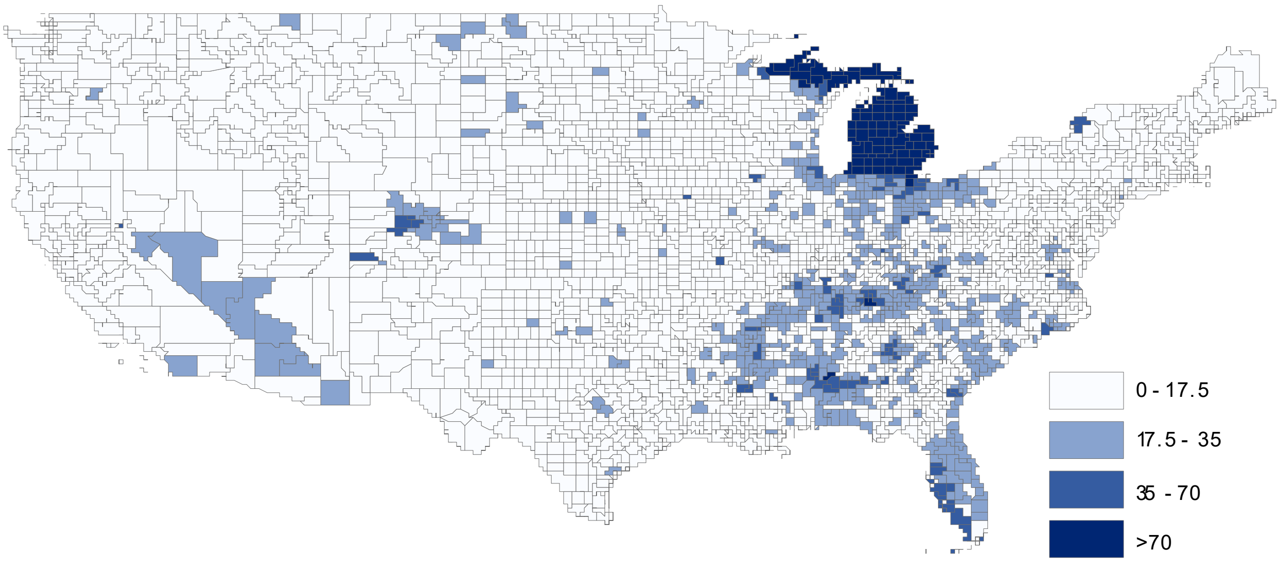
Figure A5: Connected Communities within the United States - 75 Distinct Units



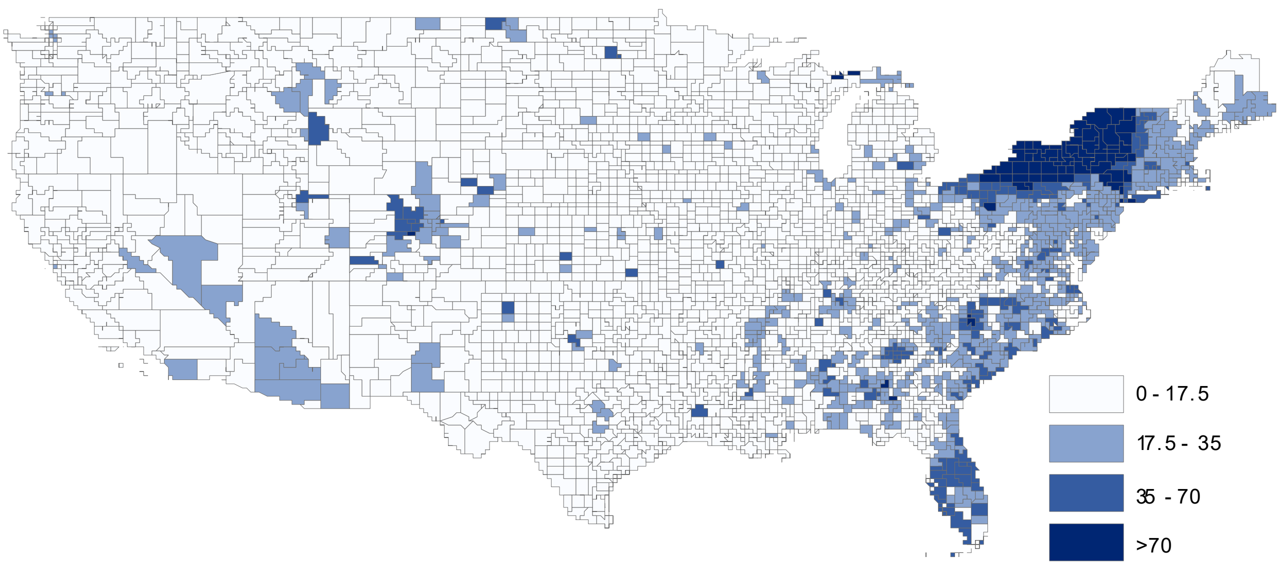
Note: Figure shows US counties grouped together when we use hierarchical agglomerative linkage clustering to create 75 distinct groups of counties. The algorithm assigns both Hawaii and Alaska, not pictured, to two distinct clusters including only the respective state.

Figure A6: State Borders and Regional Groupings

(A) Relative Probability of Friendship Link to Macomb County, MI ($RelativeProbFriendship_{i,j}$)



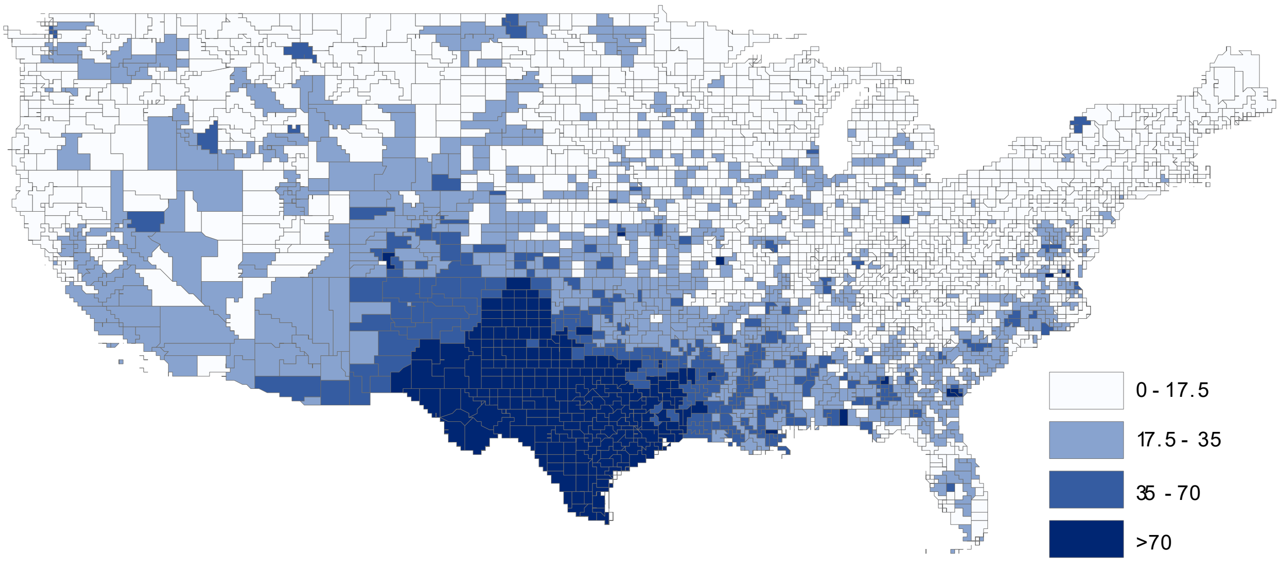
(B) Relative Probability of Friendship Link to Erie County, NY ($RelativeProbFriendship_{i,j}$)



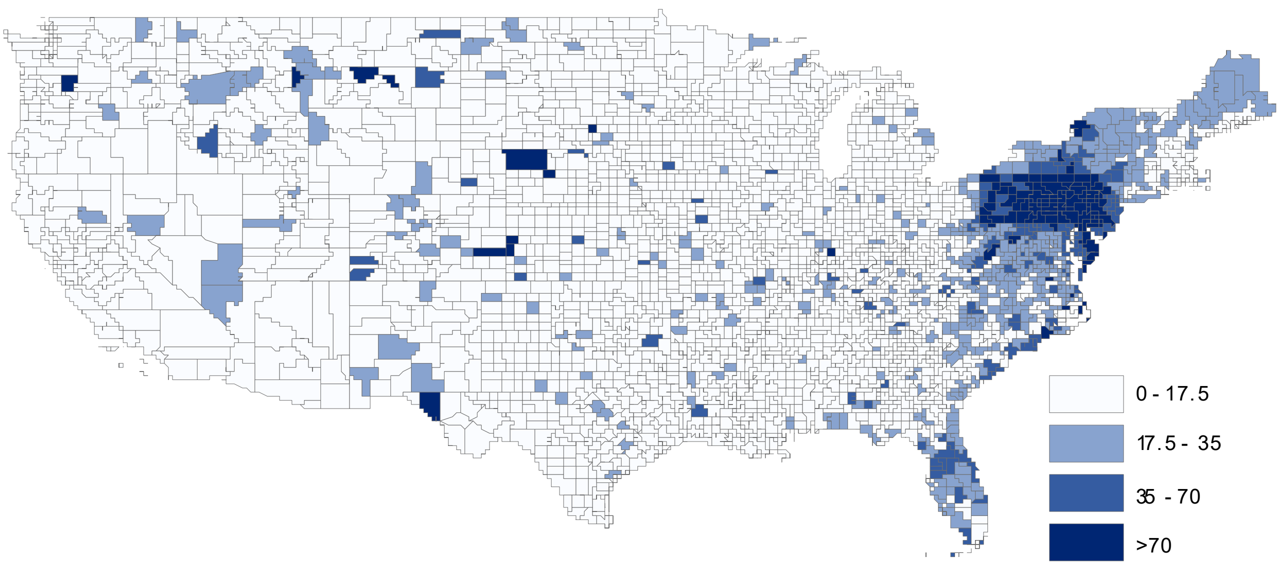
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Macomb County, MI in Panel A, and Erie County, NY in Panel B. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Macomb or Erie) and county j.

Figure A6: State Borders and Regional Groupings

(C) Relative Probability of Friendship Link to Bexar County, TX ($RelativeProbFriendship_{i,j}$)



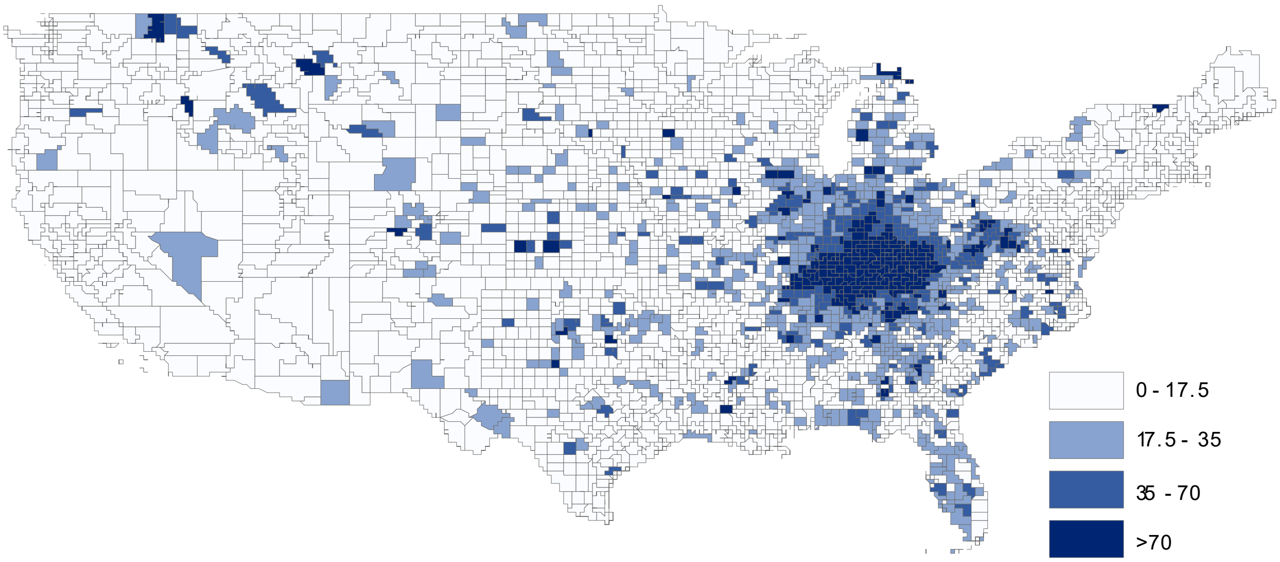
(D) Relative Probability of Friendship Link to Schuylkill County, PA ($RelativeProbFriendship_{i,j}$)



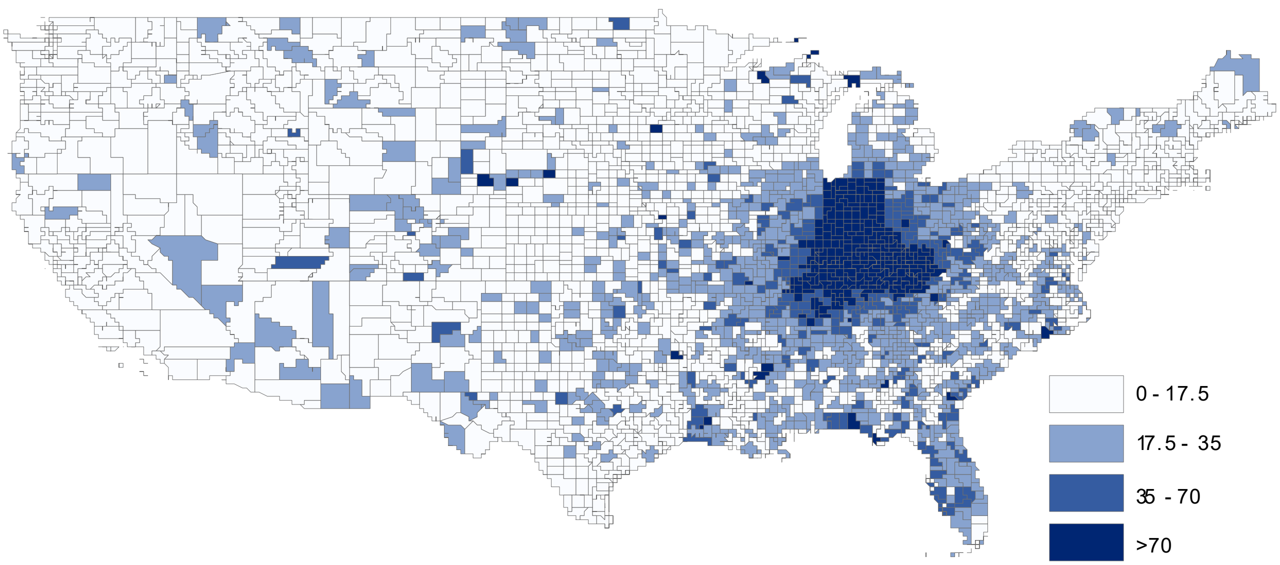
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Bexar County, TX in Panel C, and Schuylkill County, PA in Panel D. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Bexar or Schuylkill) and county j.

Figure A6: State Borders and Regional Groupings

(E) Relative Probability of Friendship Link to Marion County, KY ($RelativeProbFriendship_{i,j}$)



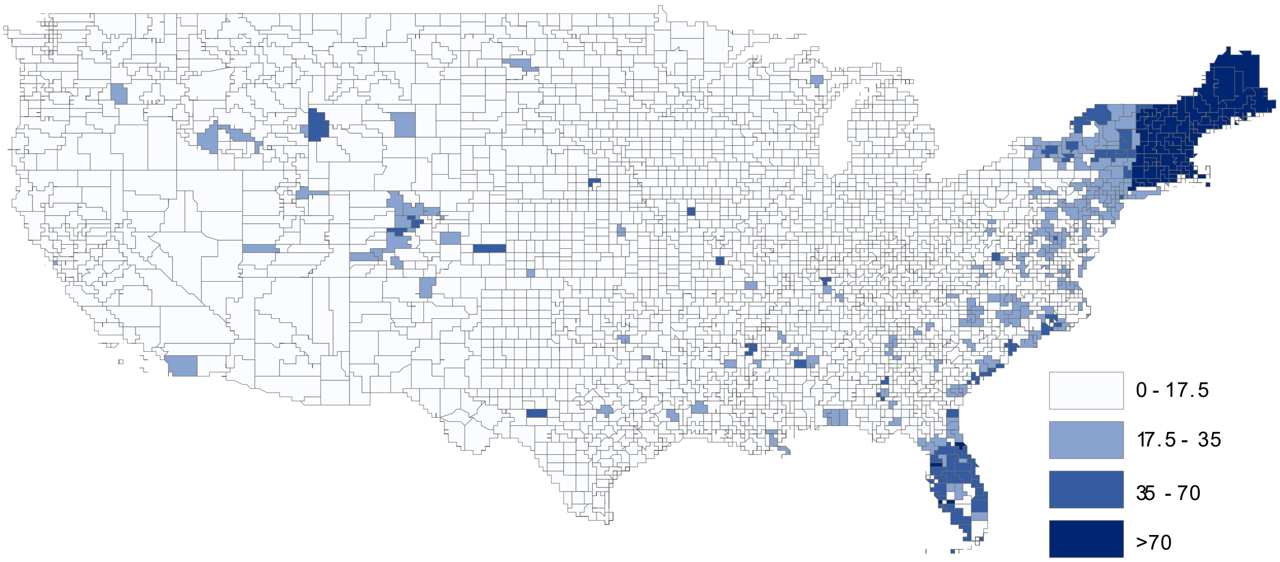
(F) Relative Probability of Friendship Link to Clark County, IN ($RelativeProbFriendship_{i,j}$)



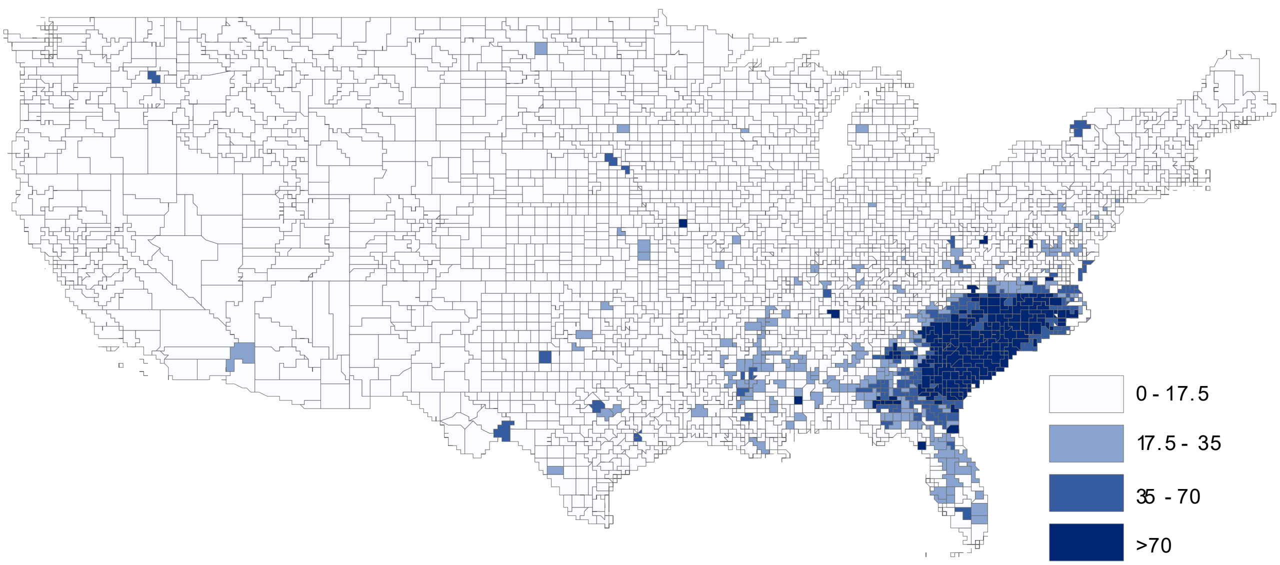
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Marion County, KY in Panel E, and Clark County, IN in Panel F. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Marion or Clark) and county j.

Figure A6: State Borders and Regional Groupings

(G) Relative Probability of Friendship Link to Bristol County, MA ($RelativeProbFriendship_{i,j}$)

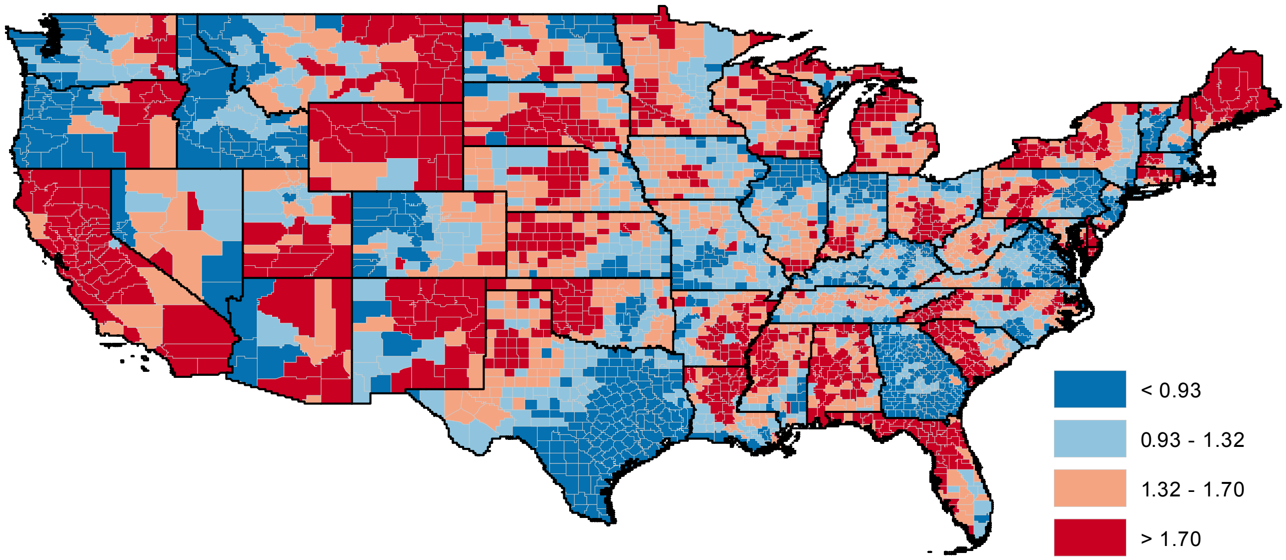


(H) Relative Probability of Friendship Link to Allendale County, SC ($RelativeProbFriendship_{i,j}$)



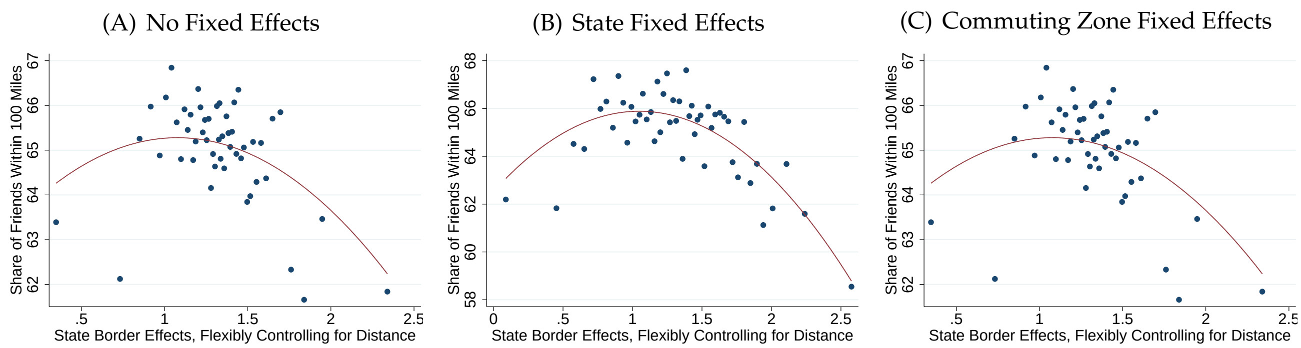
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Bristol County, MA in Panel G, and Allendale County, SC in Panel H. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Bristol or Allendale) and county j.

Figure A7: Magnitude of State Border Effects



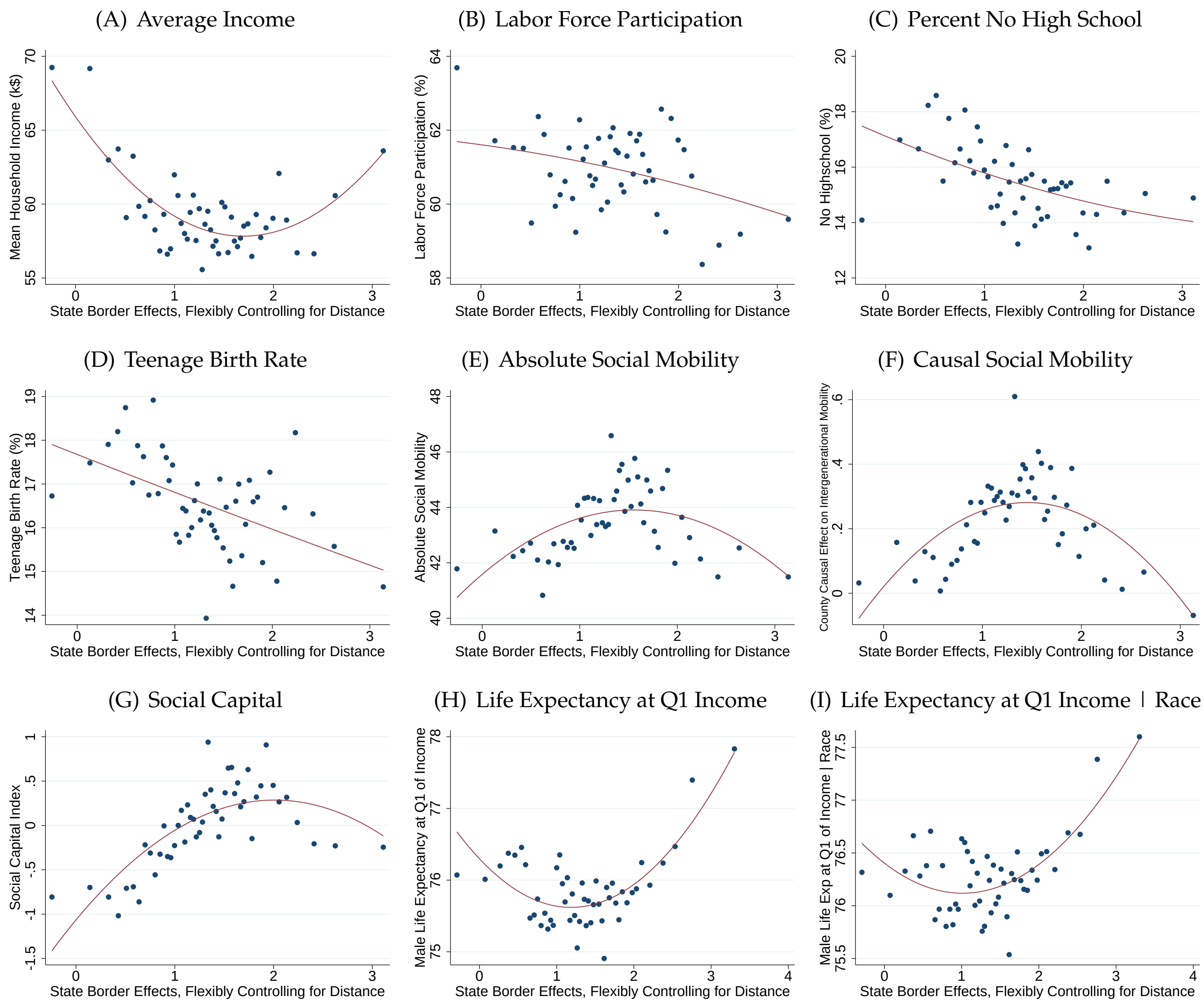
Note: This map plots the estimated state border effects by county, given as the estimated value of coefficient β_1 in Regression A3. Counties with a stronger state-border effect are red, while counties with a weaker effect are in blue.

Figure A8: State Border Effects and Share of Friends Within 100 Miles



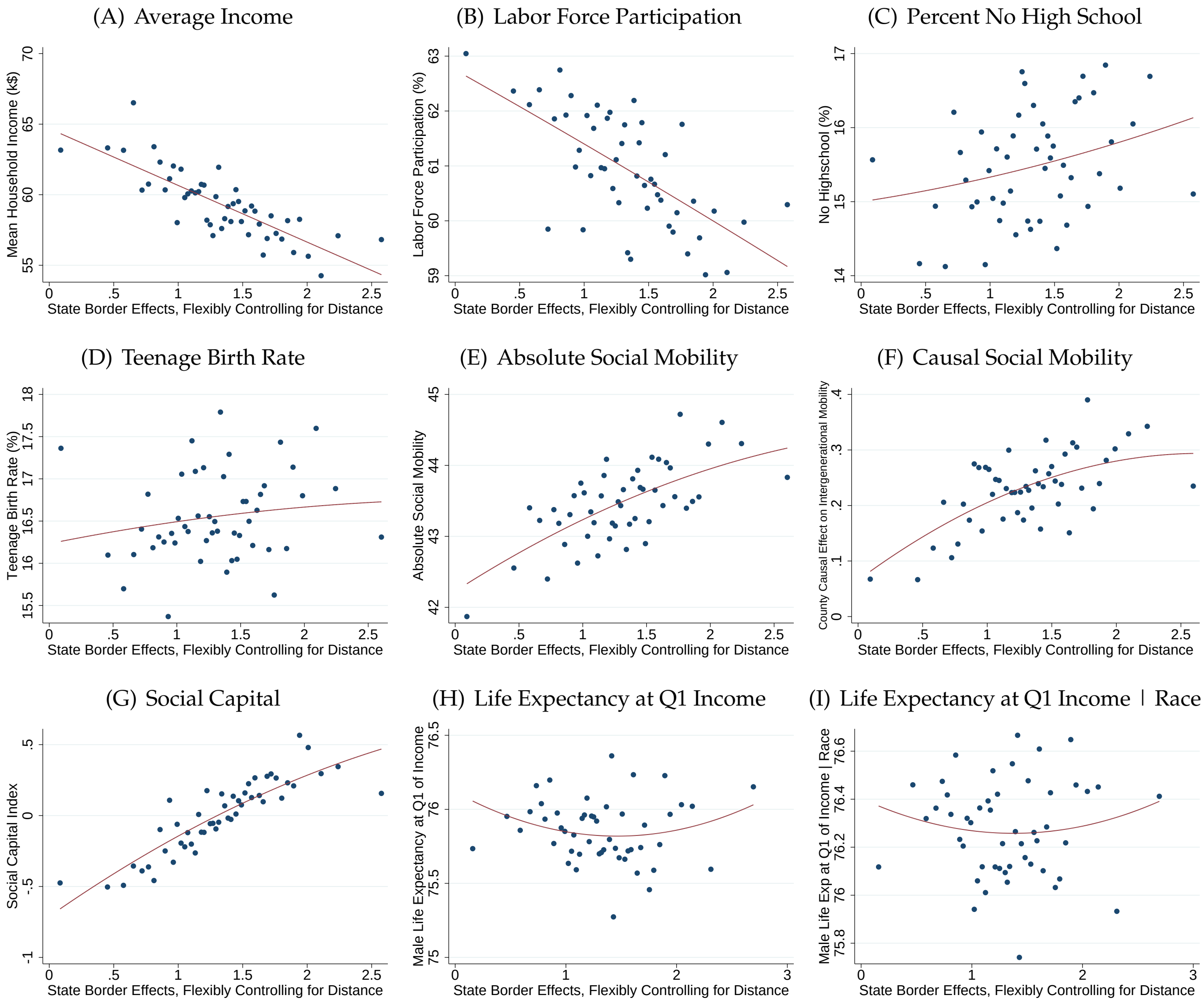
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A3. The vertical axis of each panel shows the share of friends within 100 miles for each bin. Panel A does not include fixed effects, Panel B includes state fixed effects, and Panel C includes commuting zone fixed effects.

Figure A9: State Border Effects Coefficients



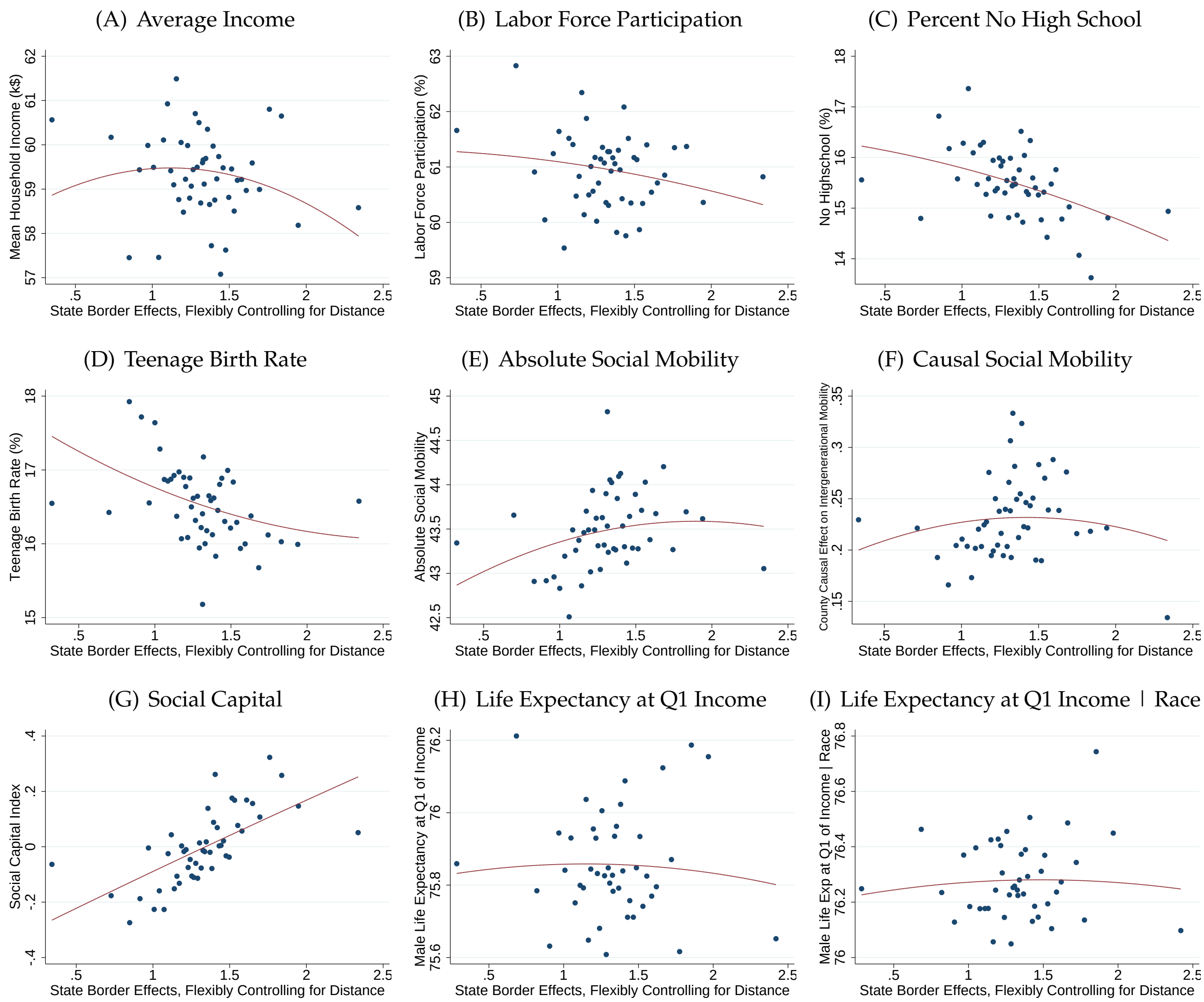
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A3. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). The red line shows the fit of a quadratic regression.

Figure A10: State Border Effects Coefficients - Conditional on State



Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A3. The regression also controls for state fixed effects. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on state fixed effects. The red line shows the fit of a quadratic regression.

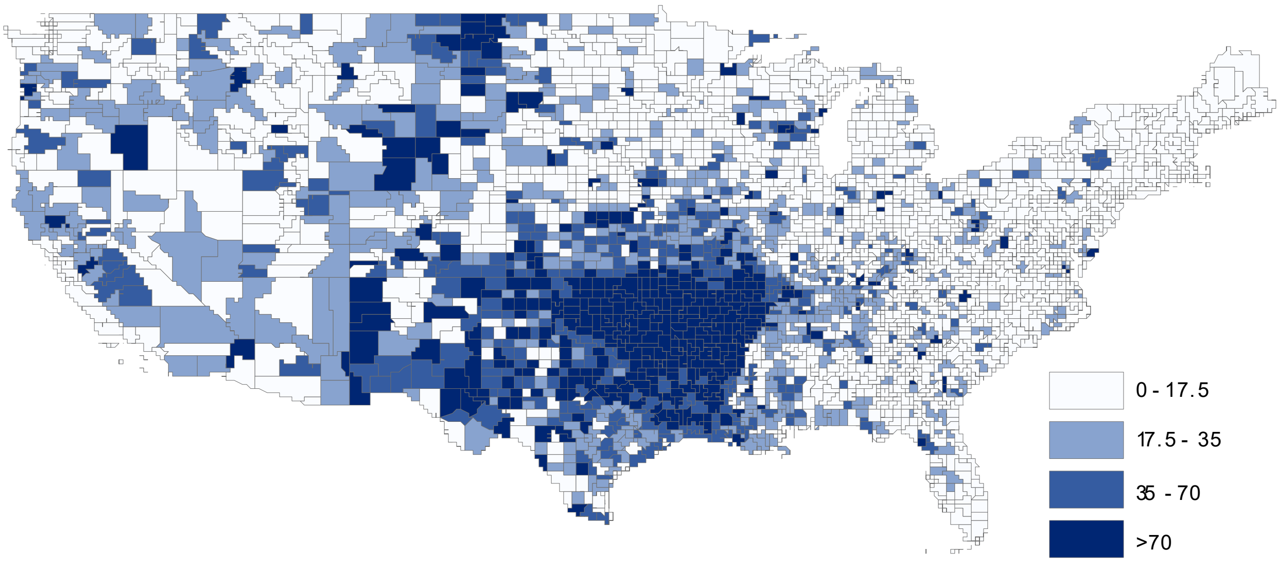
Figure A11: State Border Effects Coefficients - Conditional on Commuting Zone



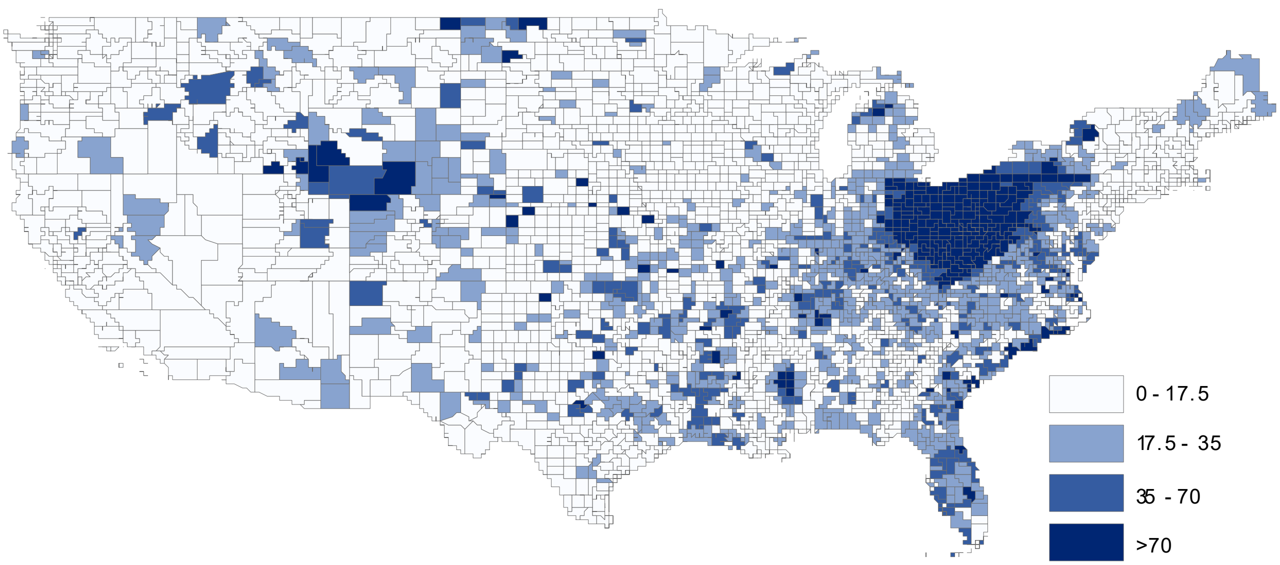
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A3. The regression also controls for commuting zone fixed effects. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on commuting zone fixed effects. The red line shows the fit of a quadratic regression.

Figure A12: Geography's Influence on Friendship Network Distribution

(A) Relative Probability of Friendship Link to Scott County, AR ($RelativeProbFriendship_{i,j}$)



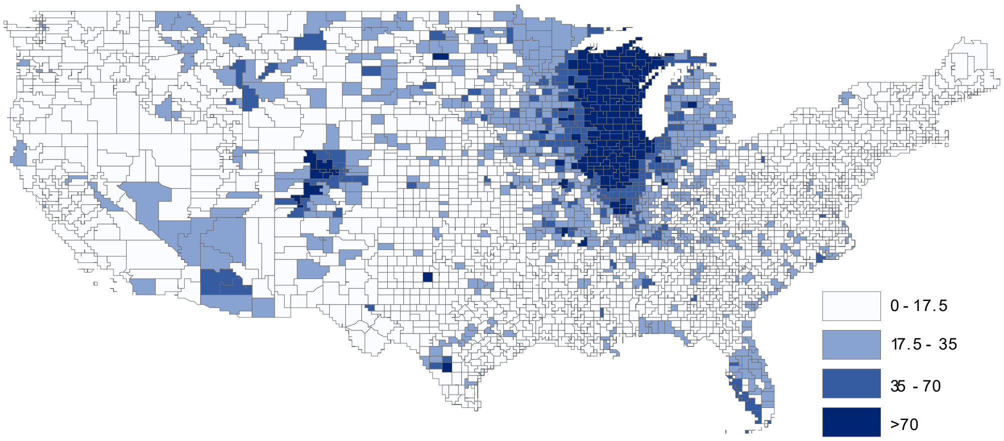
(B) Relative Probability of Friendship Link to Belmont County, OH ($RelativeProbFriendship_{i,j}$)



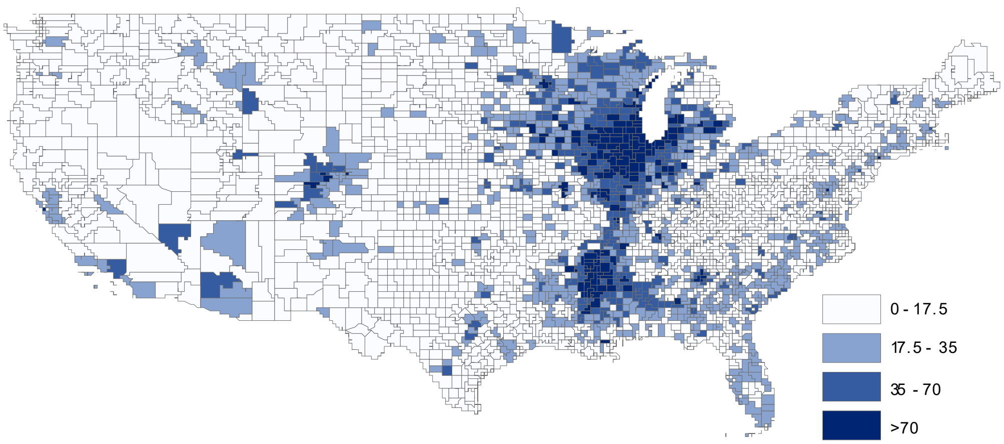
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Scott County, AR, in Panel A, and Belmont County, OH, in Panel B. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Scott or Belmont) and county j.

Figure A13: Heterogeneity in Illinois Friendship Network Distribution

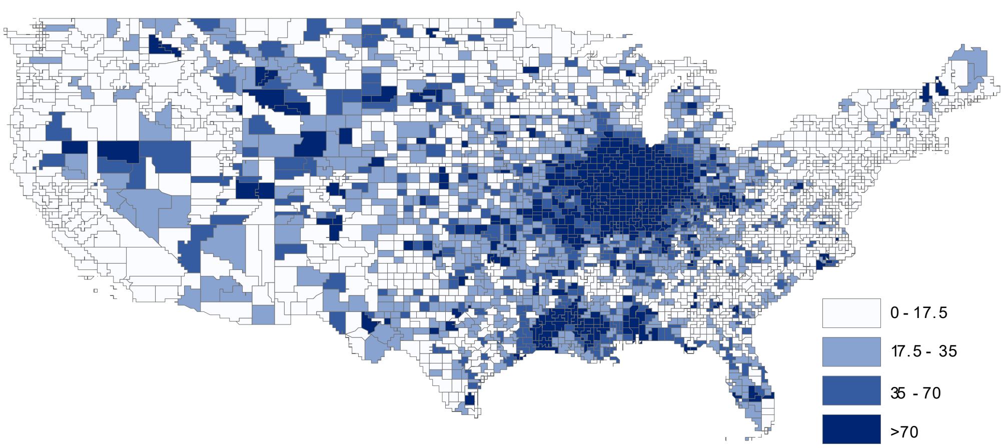
(A) Relative Probability of Friendship Link to McHenry County, IL ($RelativeProbFriendship_{i,j}$)



(B) Relative Probability of Friendship Link to Cook County, IL ($RelativeProbFriendship_{i,j}$)



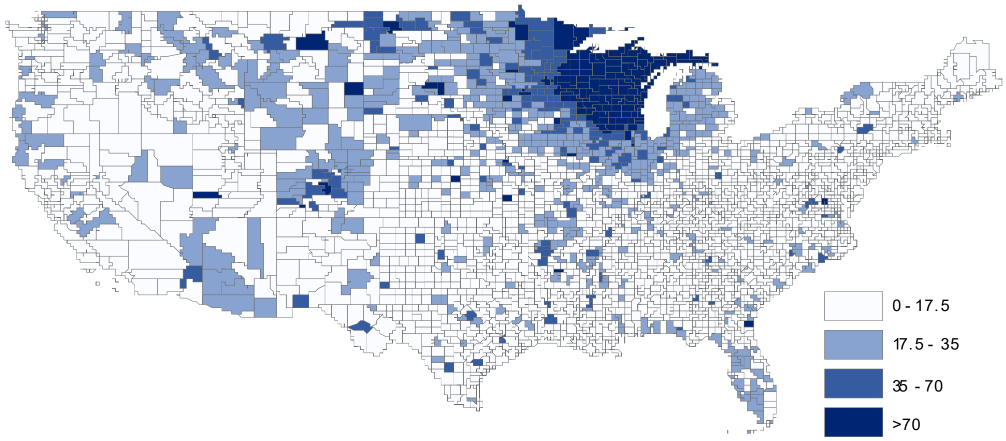
(C) Relative Probability of Friendship Link to Crawford County, IL ($RelativeProbFriendship_{i,j}$)



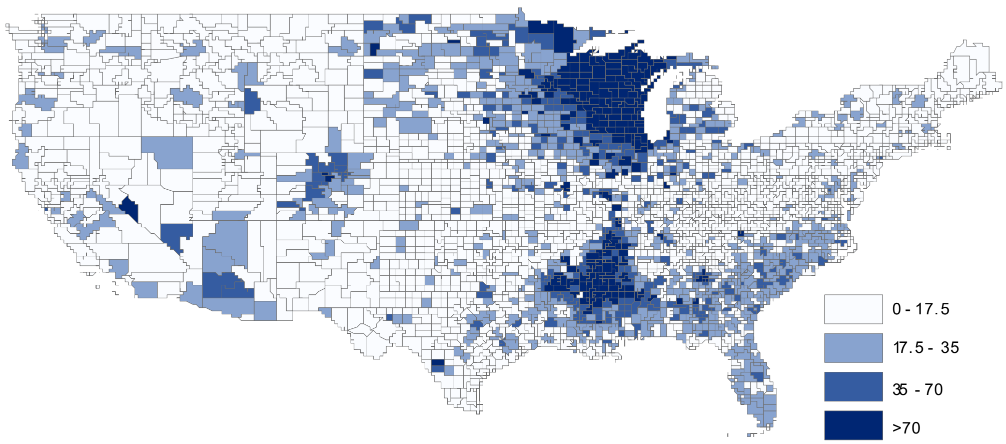
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to McHenry County, IL in Panel A, Cook County, IL in Panel B, and Crawford County, IL in Panel C. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (McHenry, Cook, or Crawford) and county j.

Figure A14: Heterogeneity in Wisconsin Friendship Network Distribution

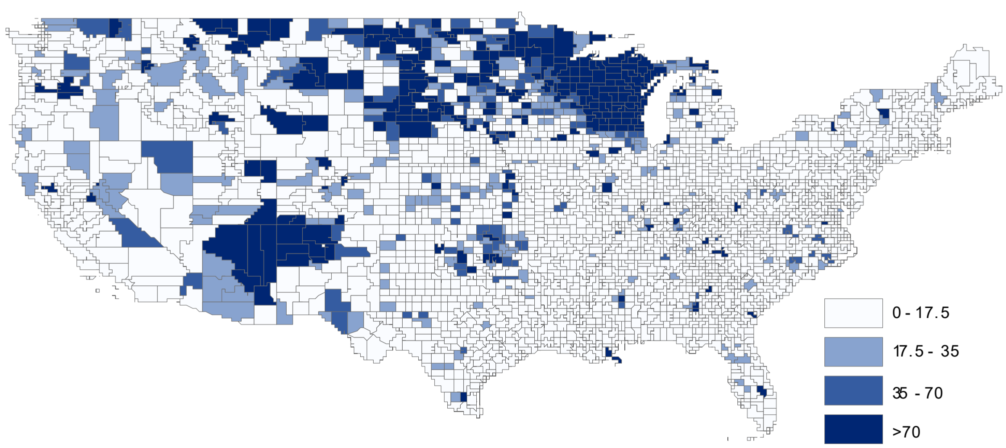
(A) Relative Probability of Friendship Link to Manitowoc County, WI ($RelativeProbFriendship_{i,j}$)



(B) Relative Probability of Friendship Link to Milwaukee County, WI ($RelativeProbFriendship_{i,j}$)



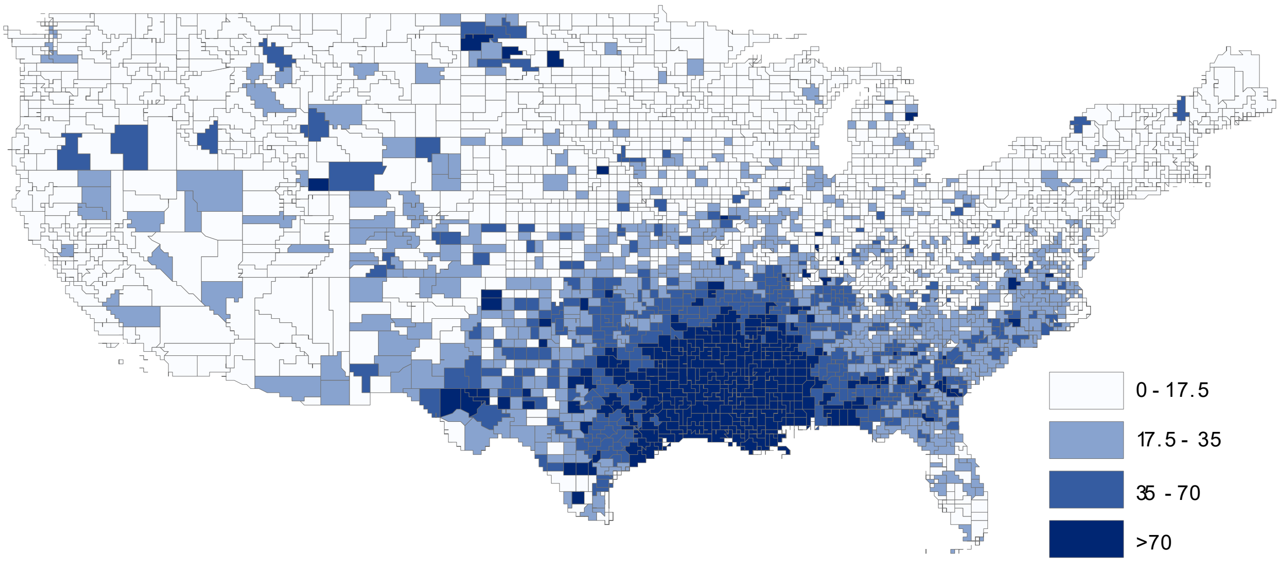
(C) Relative Probability of Friendship Link to Menominee County, WI ($RelativeProbFriendship_{i,j}$)



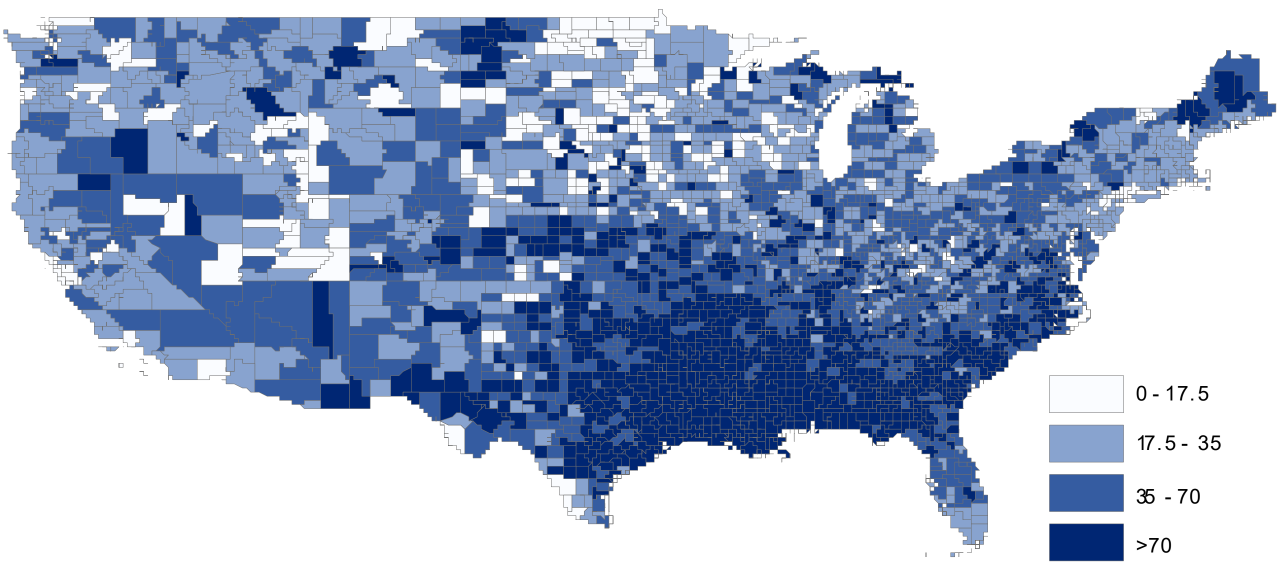
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Manitowoc County, WI in Panel A, Milwaukee County, WI in Panel B, and Menominee County, WI in Panel C. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Manitowoc, Milwaukee, or Menominee) and county j.

Figure A15: Influence of a Military Base on Friendship Network Distribution

(A) Relative Probability of Friendship Link to Rapides Parish, LA ($RelativeProbFriendship_{i,j}$)



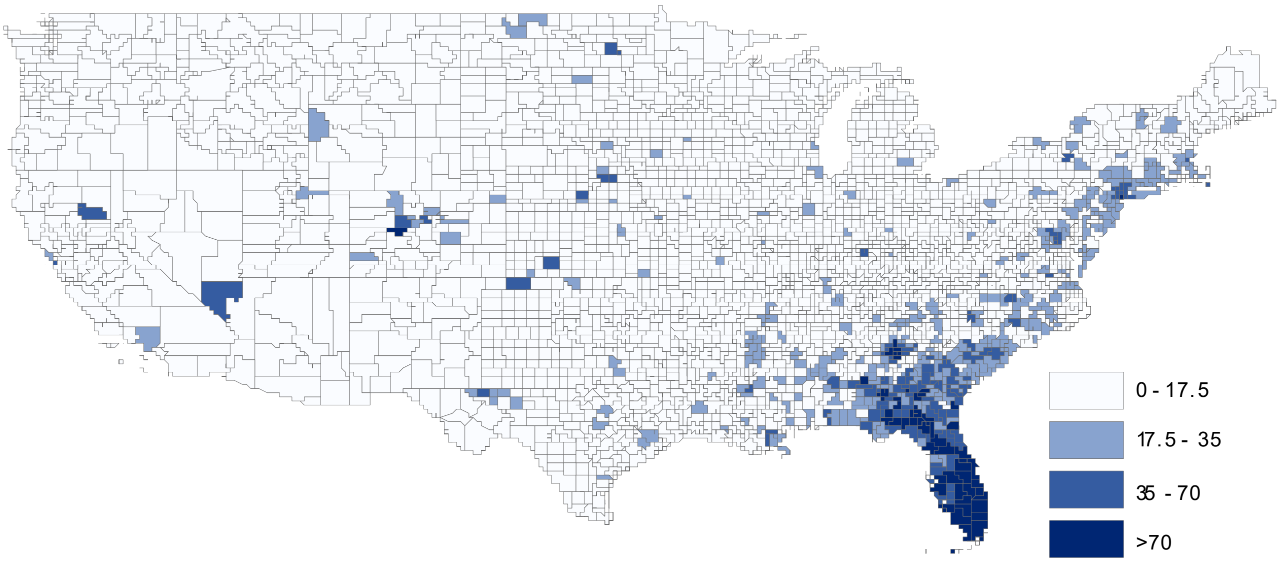
(B) Relative Probability of Friendship Link to Vernon Parish, LA ($RelativeProbFriendship_{i,j}$)



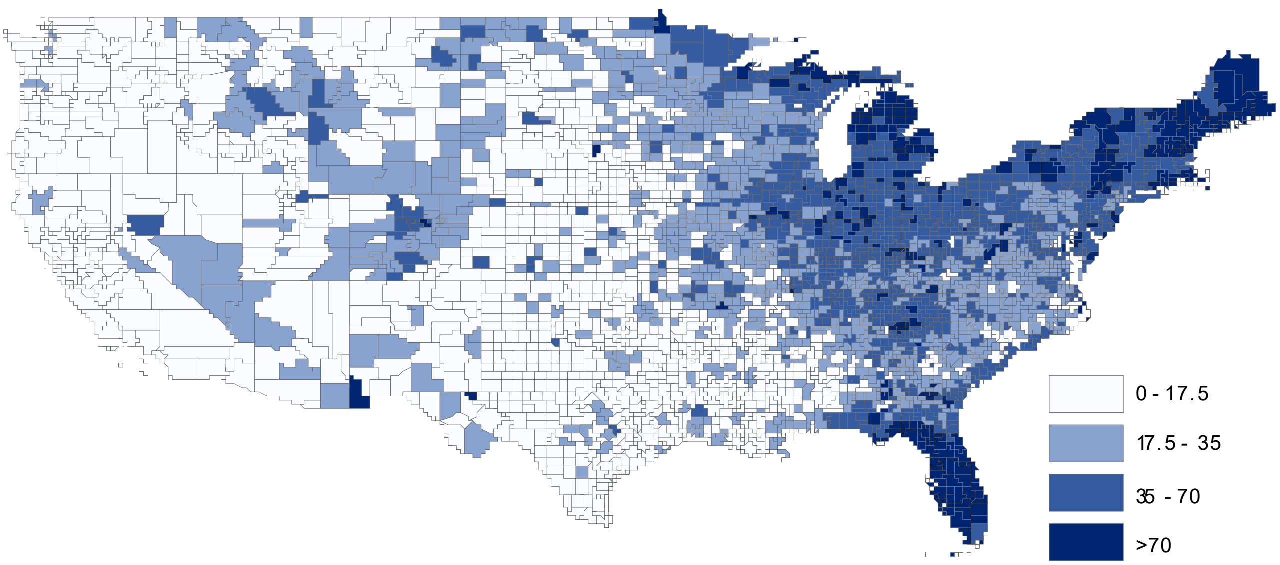
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Rapides Parish, LA in Panel G, and Vernon Parish, LA in Panel H. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Rapides or Vernon) and county j.

Figure A16: Florida Retirement Communities and Friendship Network Distribution

(A) Relative Probability of Friendship Link to Miami-Dade County, FL ($RelativeProbFriendship_{i,j}$)



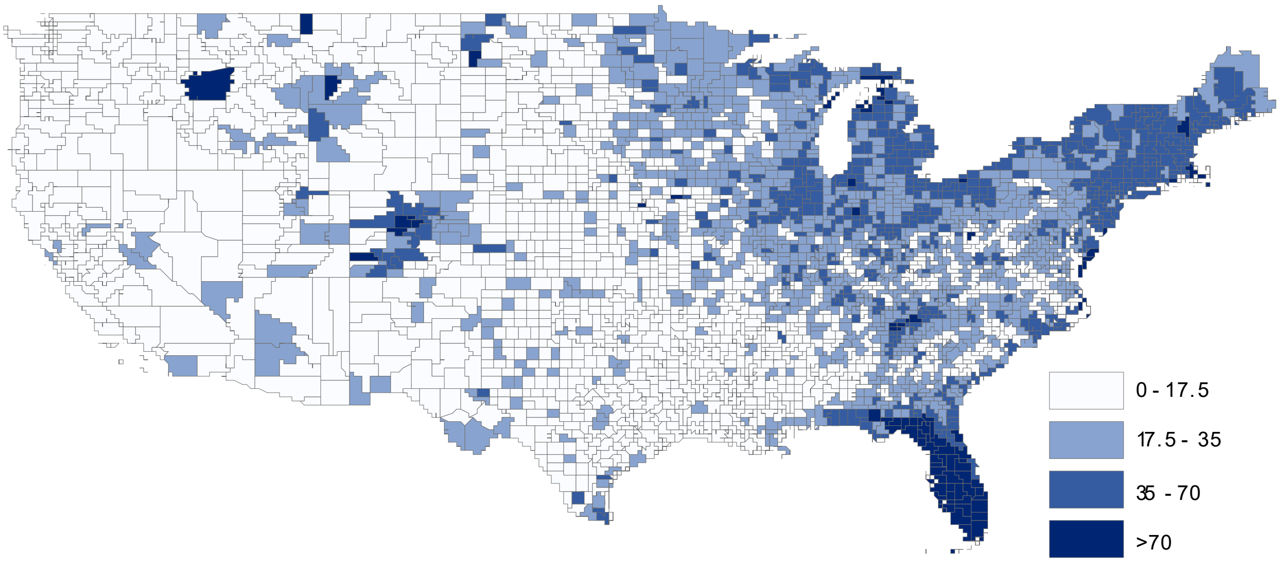
(B) Relative Probability of Friendship Link to Charlotte County, FL ($RelativeProbFriendship_{i,j}$)



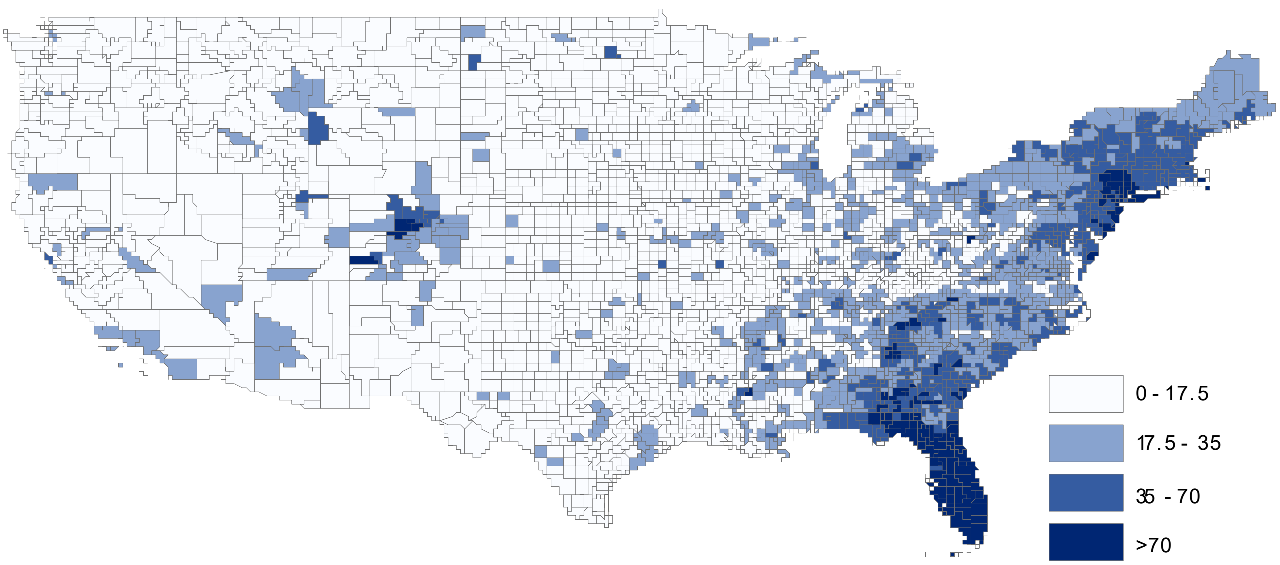
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Miami-Dade County, FL in Panel A, and Charlotte County, FL in Panel B. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Miami-Dade or Charlotte) and county j.

Figure A16: Florida Retirement Communities and Friendship Network Distribution

(C) Relative Probability of Friendship Link to Collier County, FL ($RelativeProbFriendship_{i,j}$)



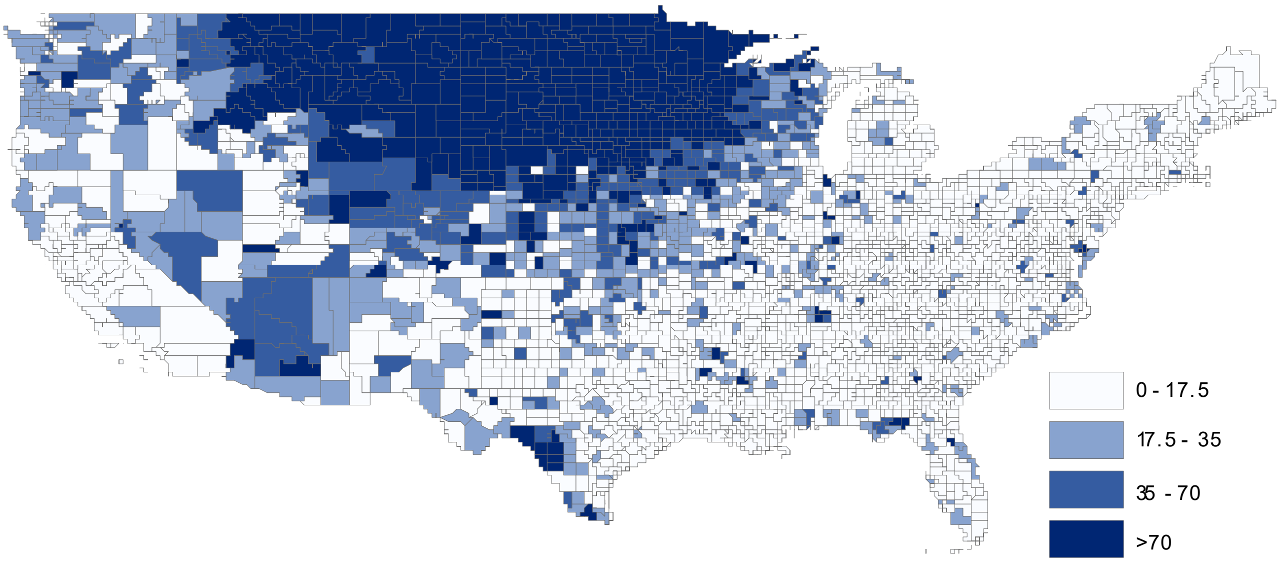
(D) Relative Probability of Friendship Link to Palm Beach County, FL ($RelativeProbFriendship_{i,j}$)



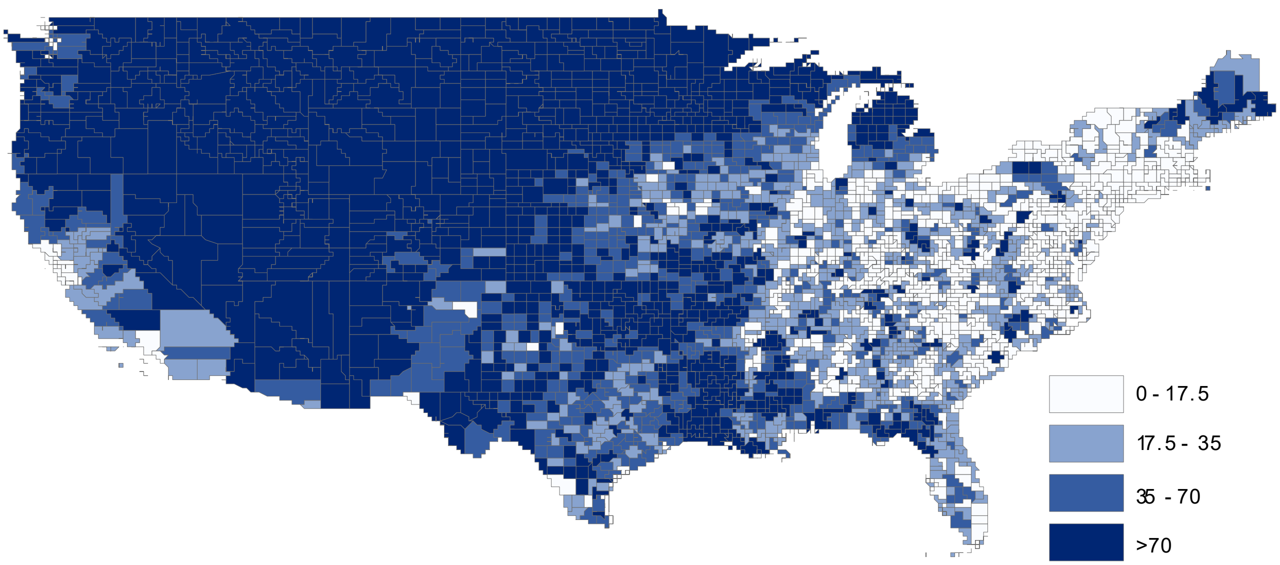
Note: Figure shows the relative probability that a Facebook user in each county j has a friendship link to Collier County, FL in Panel C, and Palm Beach County, FL in Panel D. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Collier or Palm Beach) and county j.

Figure A17: North Dakota Shale Oil Boom and Friendship Network Distribution

(A) Relative Probability of Friendship Link to Richlands County, ND ($RelativeProbFriendship_{i,j}$)



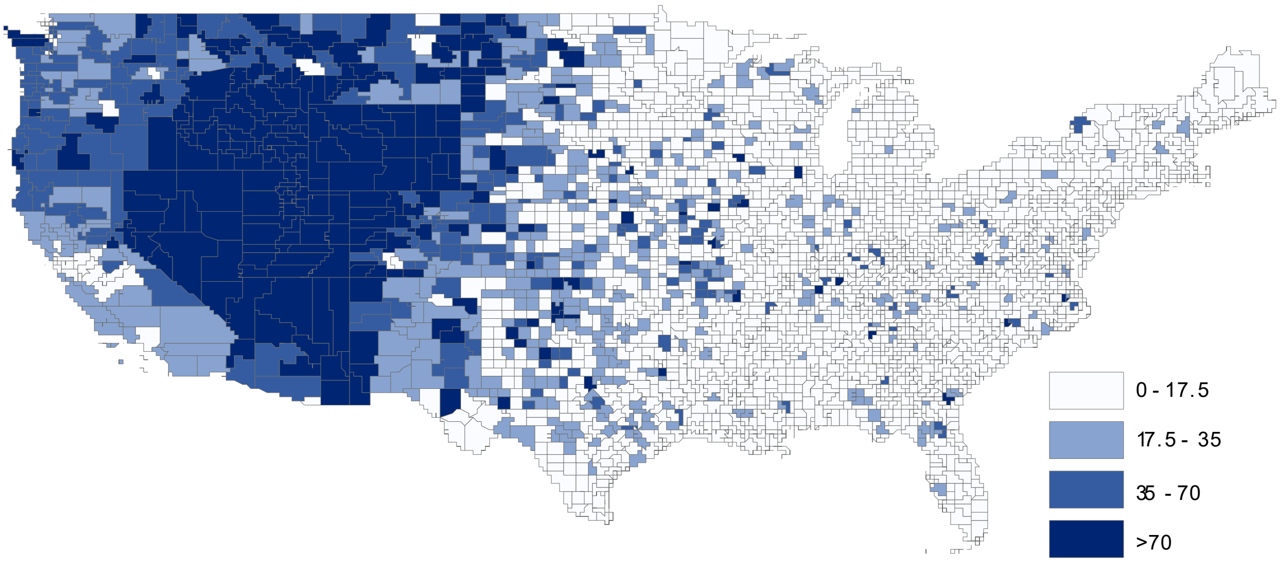
(B) Relative Probability of Friendship Link to McKenzie County, ND ($RelativeProbFriendship_{i,j}$)



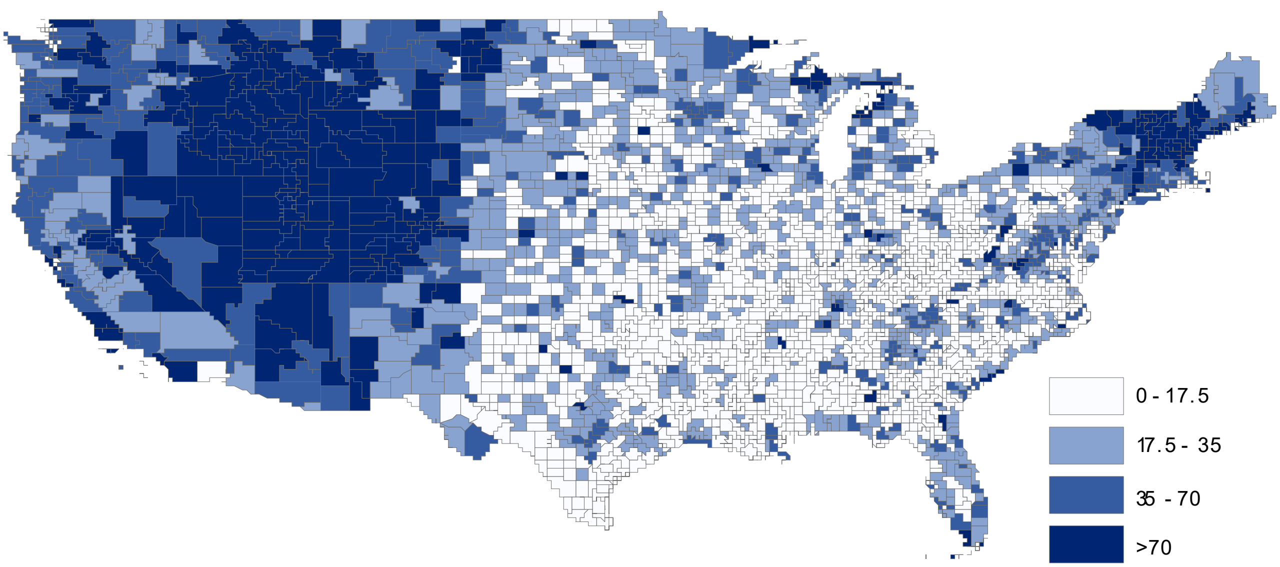
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Richlands County, ND in Panel A, and McKenzie County, ND in Panel B. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Richlands or McKenzie) and county j.

Figure A18: Linkages Between Geographically Distant Winter Sports Areas

(A) Relative Probability of Friendship Link to Sanpete County, UT ($RelativeProbFriendship_{i,j}$)

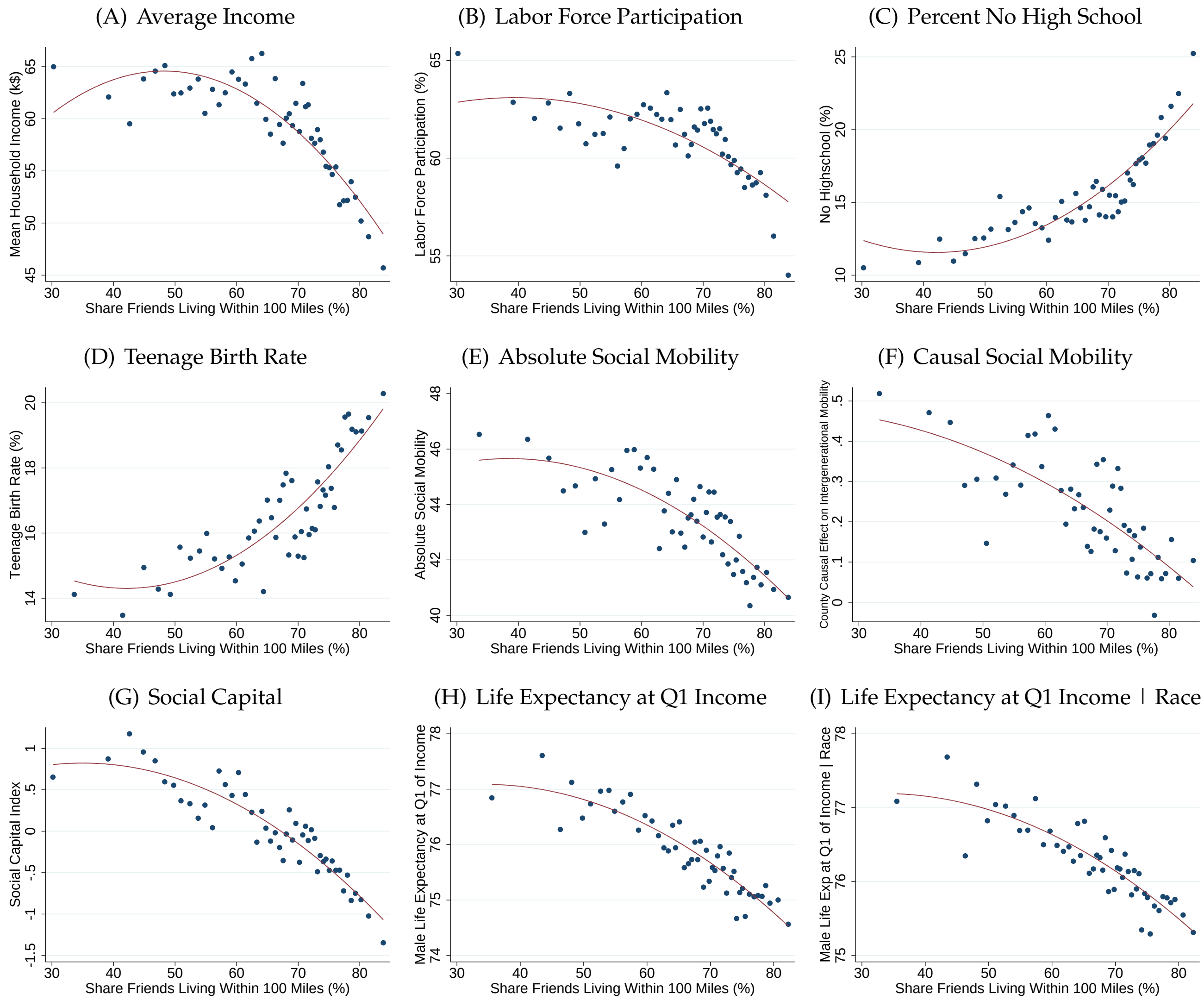


(B) Relative Probability of Friendship Link to Summit County, UT ($RelativeProbFriendship_{i,j}$)



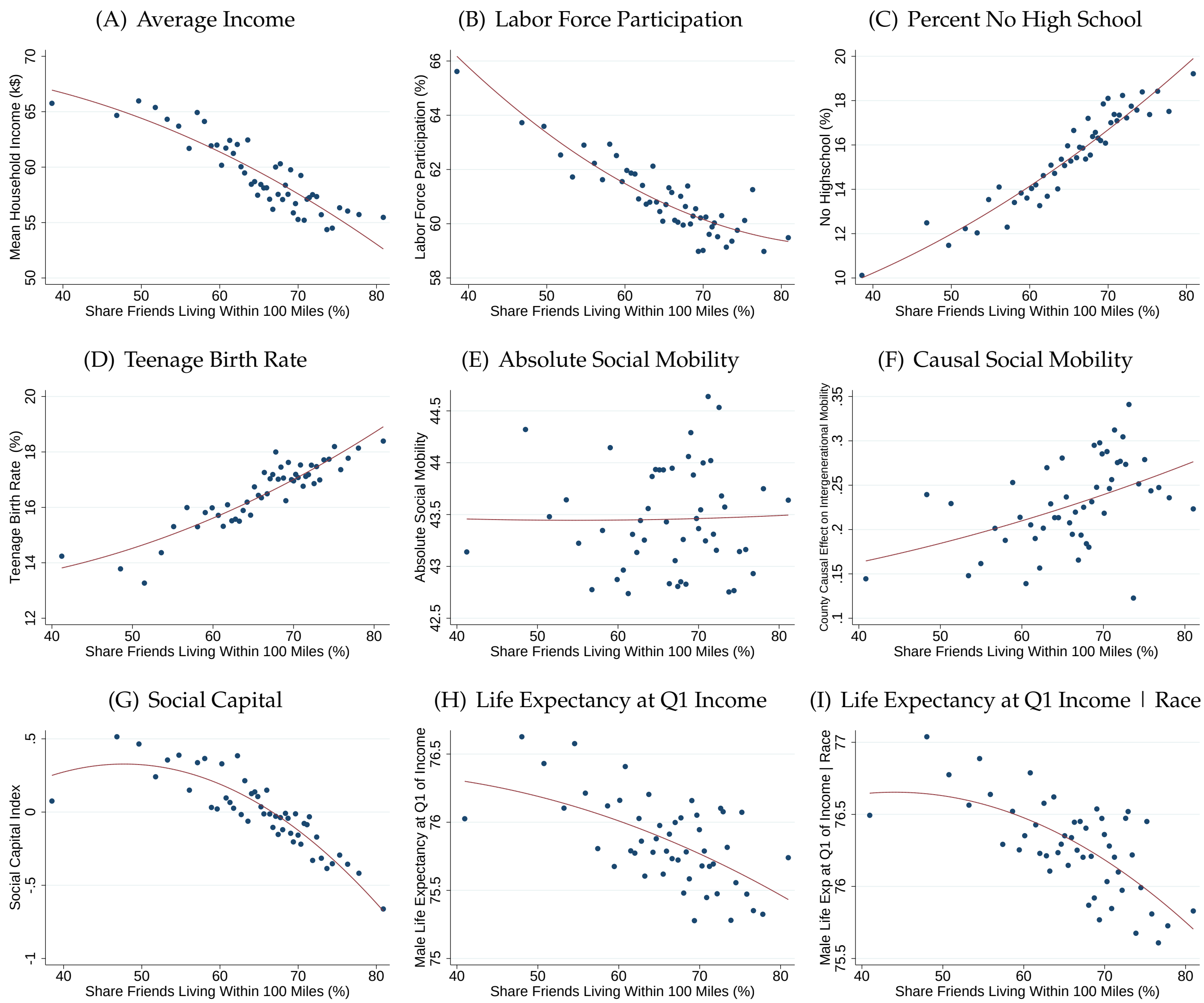
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Sanpete County, UT in Panel A, and Summit County, UT in Panel B. It is constructed as in equation A1. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Sanpete or Summit) and county j.

Figure A19: Share of Friends Within 100 Miles



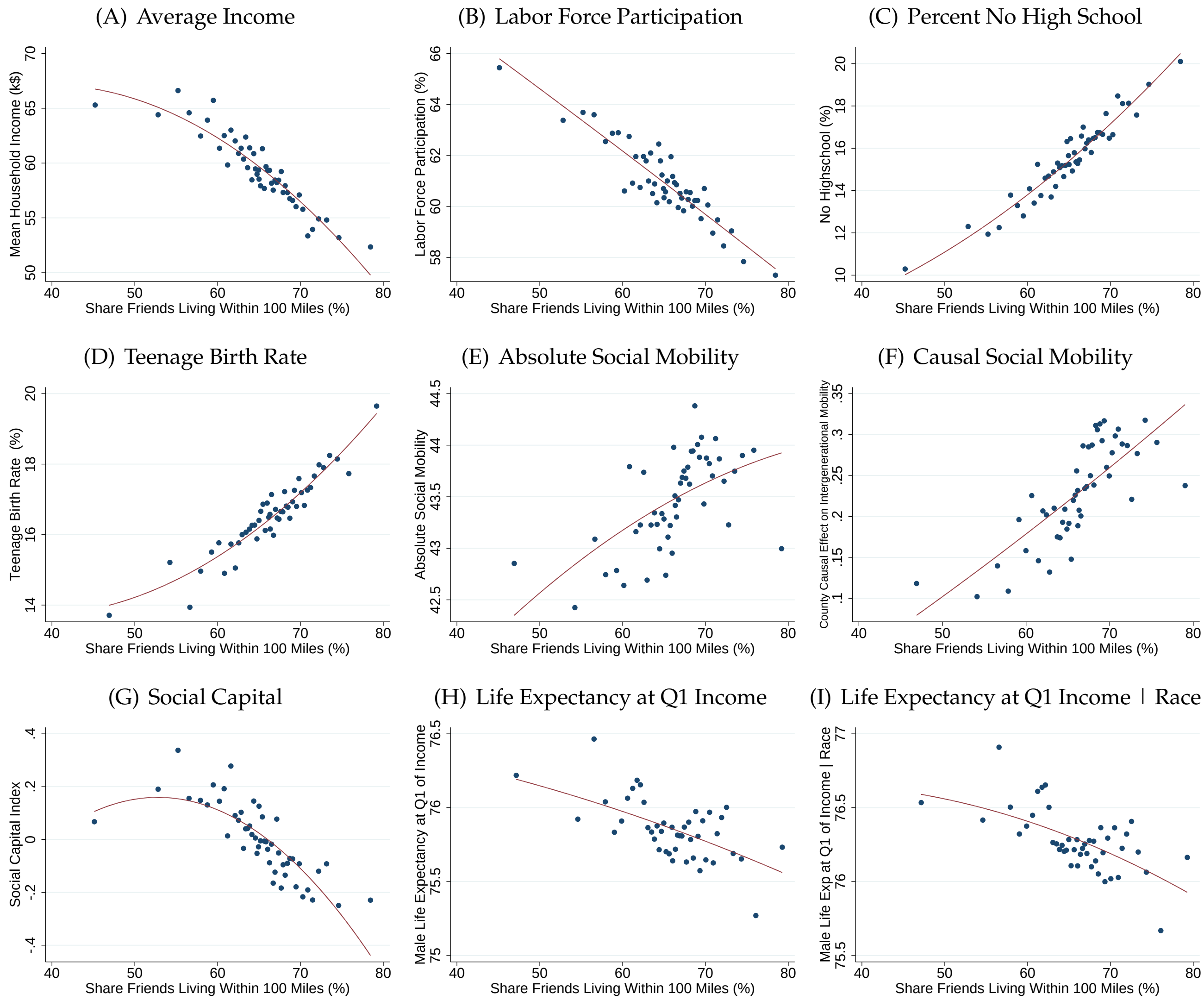
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). The red line shows the fit of a quadratic regression.

Figure A20: Share of Friends Within 100 Miles - Conditional on State



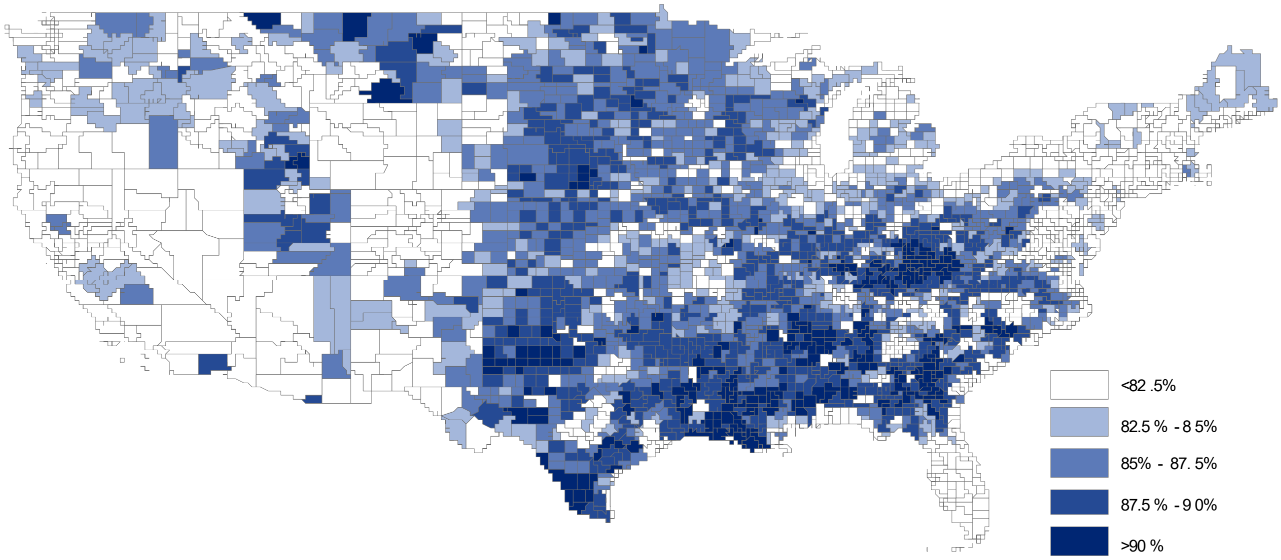
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on state fixed effects. The red line shows the fit of a quadratic regression.

Figure A21: Share of Friends Within 100 Miles - Conditional on Commuting Zone



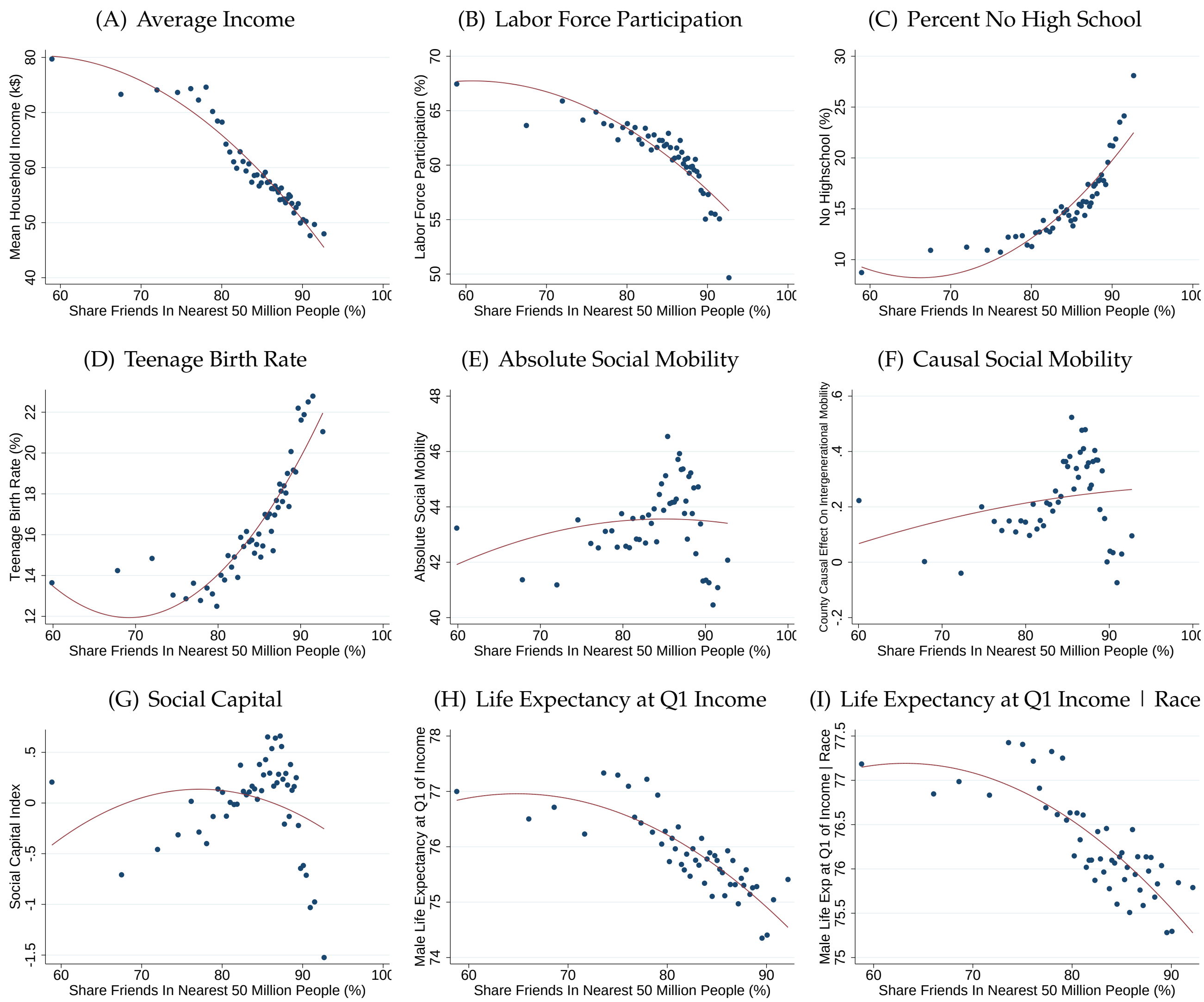
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on commuting zone fixed effects. The red line shows the fit of a quadratic regression.

Figure A22: Share of Friends Among Nearest 50 Million People



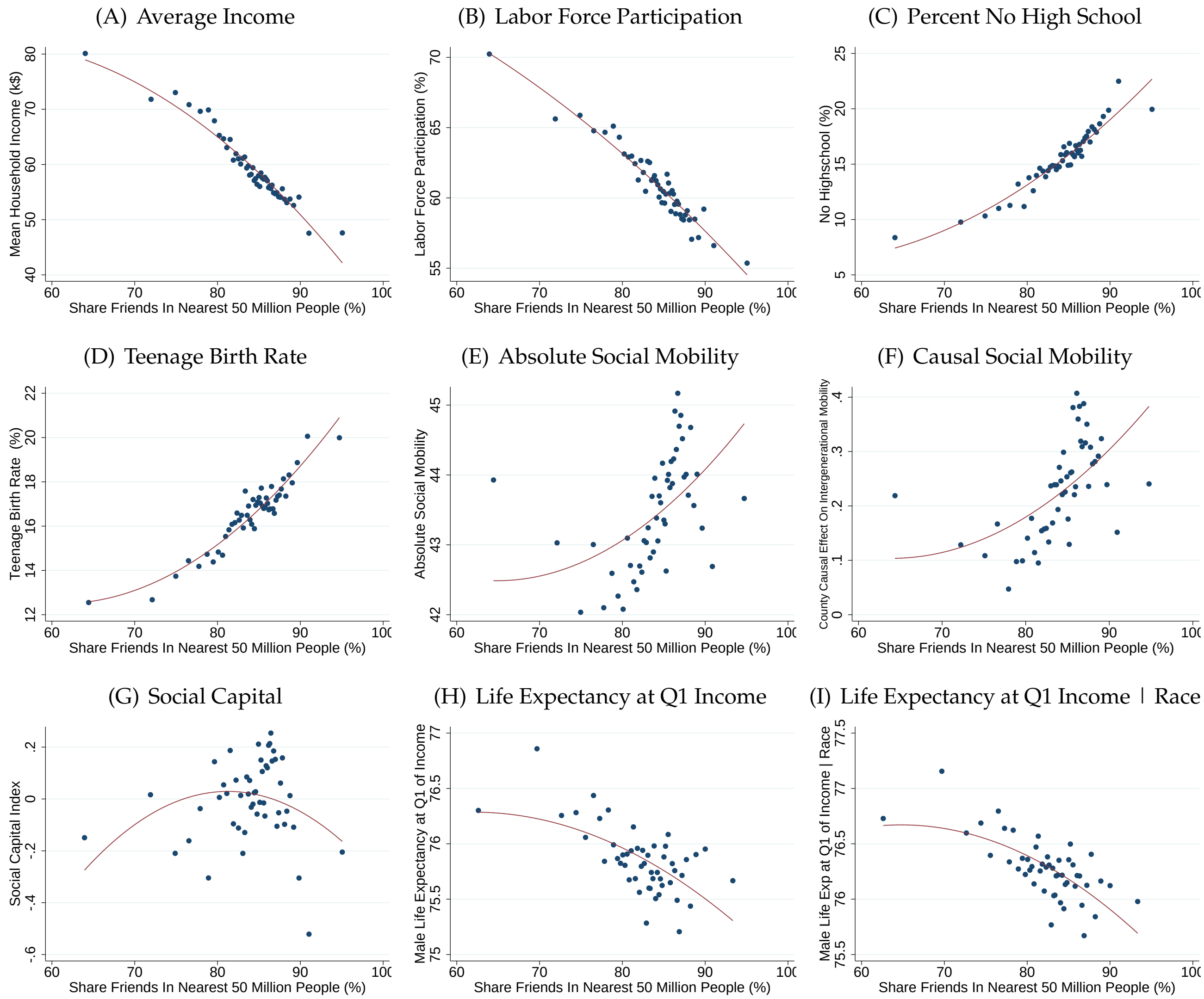
Note: Figure shows a map at the county level of the share of all US friends that are among the nearest 50 million people.

Figure A23: Share of Friends Among Nearest 50 Million People



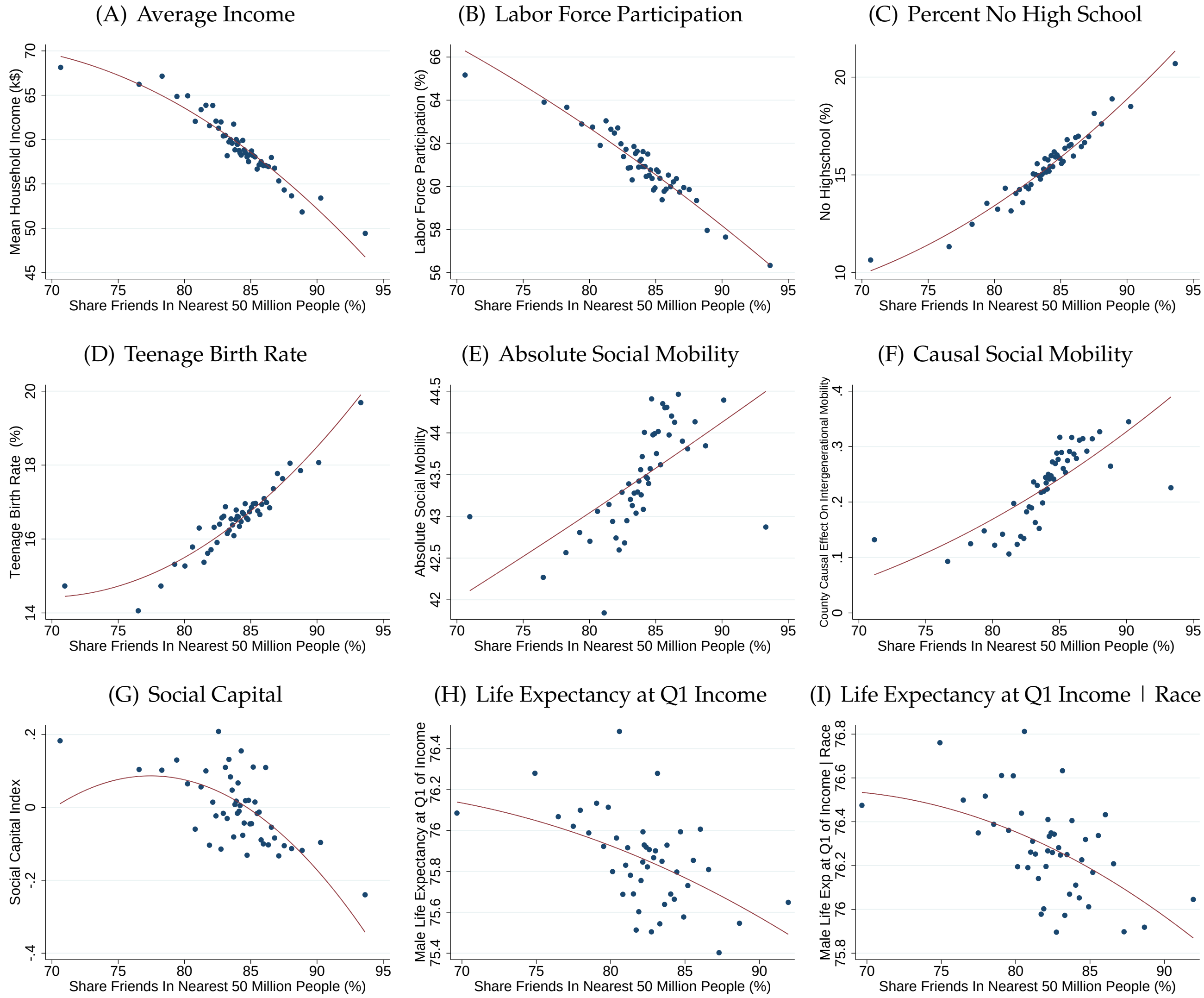
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within the nearest 50 million people. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). The red line shows the fit of a quadratic regression.

Figure A24: Share of Friends Among Nearest 50 Million People - Conditional on State



Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within the nearest 50 million people. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on state fixed effects. The red line shows the fit of a quadratic regression.

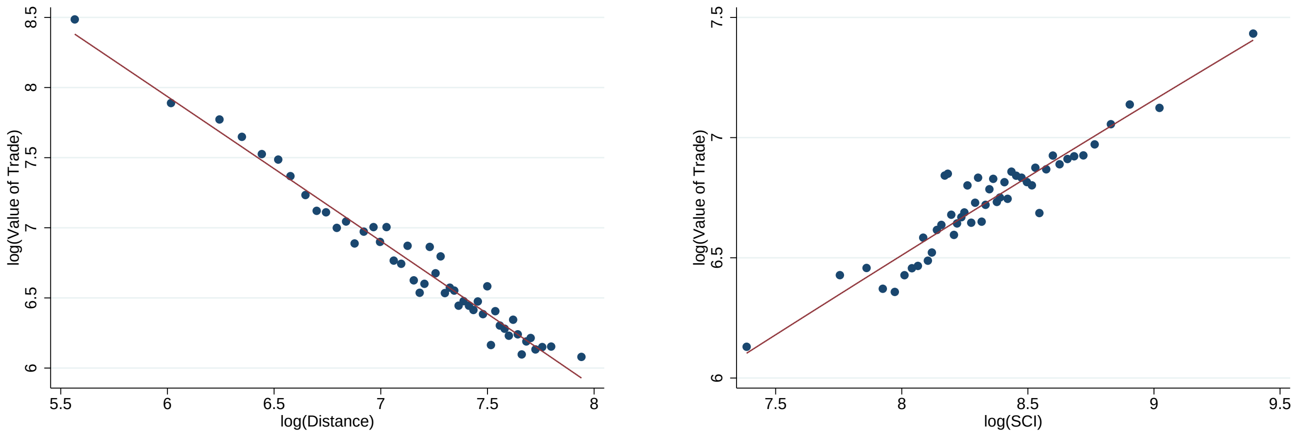
Figure A25: Share of Friends Among Nearest 50 Million People - Conditional on Commuting Zone



Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within the nearest 50 million people. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force participation (Panel B), the share of the population with no high school degree (Panel C), the teenage birth rate as provided by Chetty et al. (2014) in Panel D, the absolute measure of social mobility from Chetty et al. (2014) in Panel E, the causal measure of social mobility from Chetty and Hendren (2015) in Panel F, the measure of social capital in 2009 as defined by Rupasingha, Goetz and Freshwater (2006) in Panel G, and the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016), both unconditional (Panel H) and conditional on race (Panel I). All panels show results conditional on commuting zone fixed effects. The red line shows the fit of a quadratic regression.

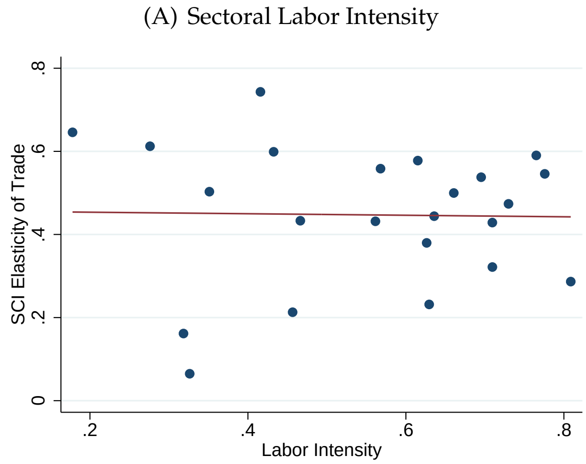
Figure A26: State-Level Trade Flows

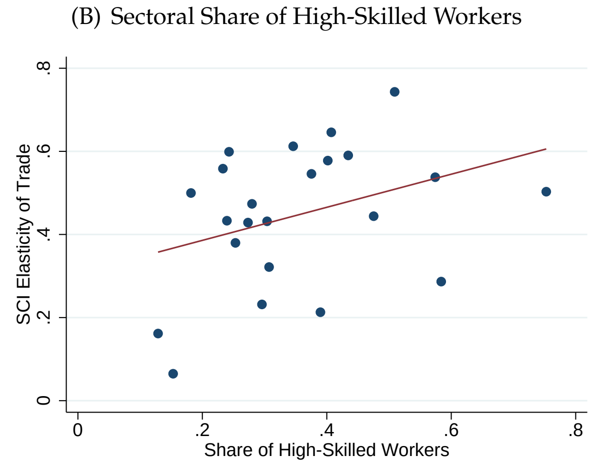
(A) Geographic Distance (B) Friendship Links



Note: Both panels show scatter plots at the state-pair level, and have the log of the trade flow between these states on the vertical axis. In Panel A, the log of the geographic distances between the states is on the horizontal axis, and in Panel B, the log of the SCI is on the horizontal axis. Both panels control for state fixed effects, and include dummies for within-state flows, and for flows to neighboring states. Panel B also controls flexibly for the log of the geographic distance between the states.

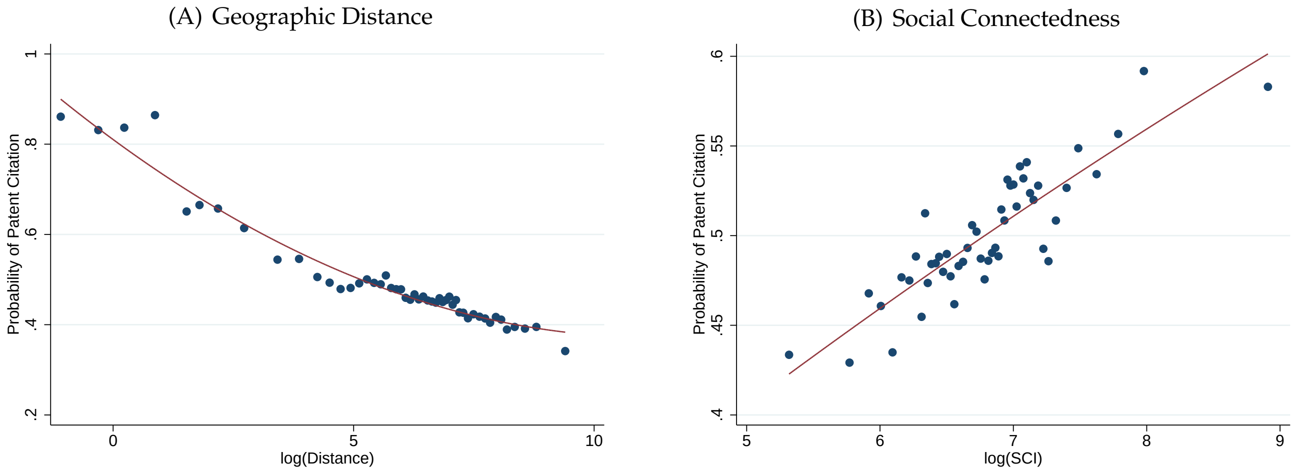
Figure A27: Sectoral Friendship Elasticities of Trade





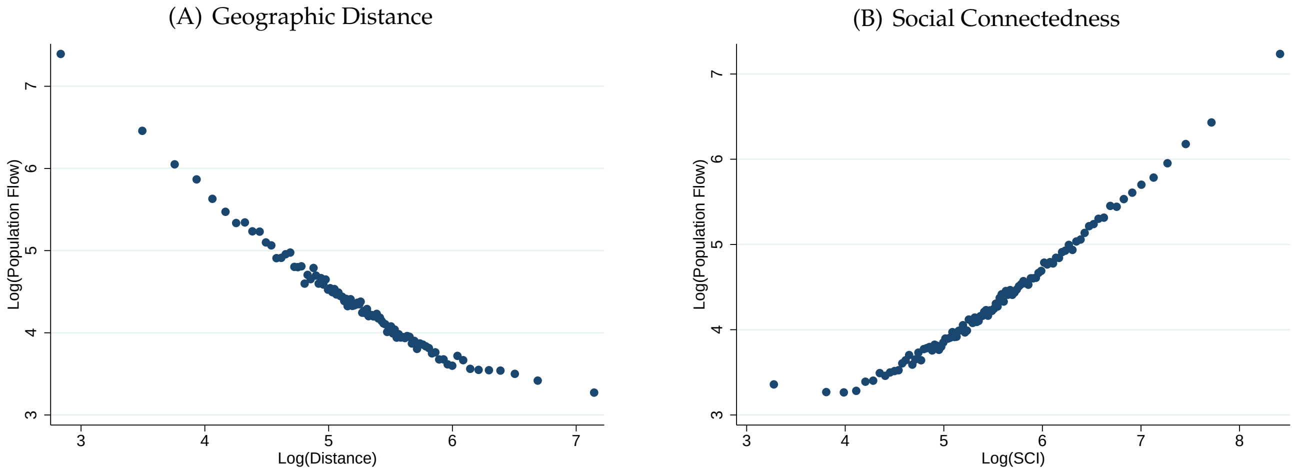
Note: Both panels show binned scatter plots at the sectoral level, with the friendship elasticities of trade for each sector on the vertical axis. In Panel A, the labor intensity of the sector is on the horizontal axis, and in Panel B, the share of high-skilled workers in the sector is on the horizontal axis.

Figure A28: Patent Citations



Note: Both panels show binned scatter plots of the probability of a patent citation on the vertical axis. Panel A plots the log of distance between the counties of the issued and cited patents on the horizontal axis. Panel B plots the log of the SCI between the counties of the issued and cited patents on the horizontal axis. Both plots control for patent class and county fixed effects, and Panel B also controls flexibly for the log of the geographic distance between counties.

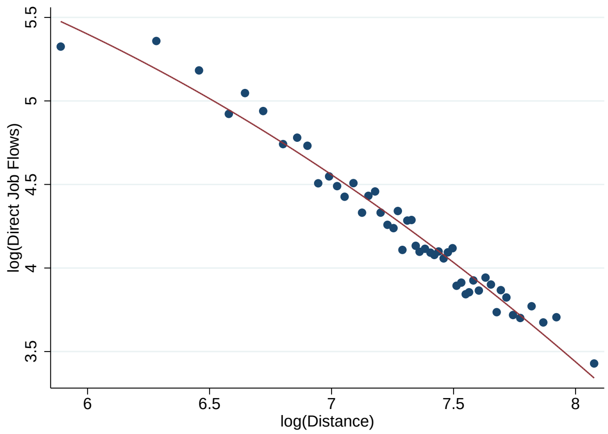
Figure A29: County-Level Migration



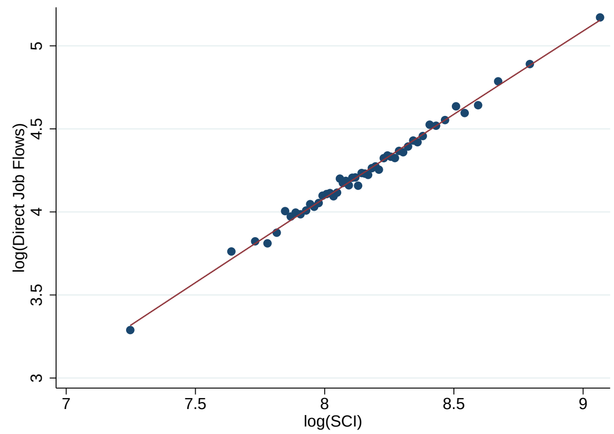
Note: Both panels show binned scatter plots at the county-pair level, and plot the log of the migration flow between these counties on the vertical axis. In Panel A, the log of the geographic distances between the counties is on the horizontal axis, and in Panel B, the log of the SCI is on the horizontal axis. Both panels control for county fixed effects, and Panel B also controls flexibly for the log of the geographic distance between the counties.

Figure A30: State-Level Labor Flows

(A) Geographic Distance



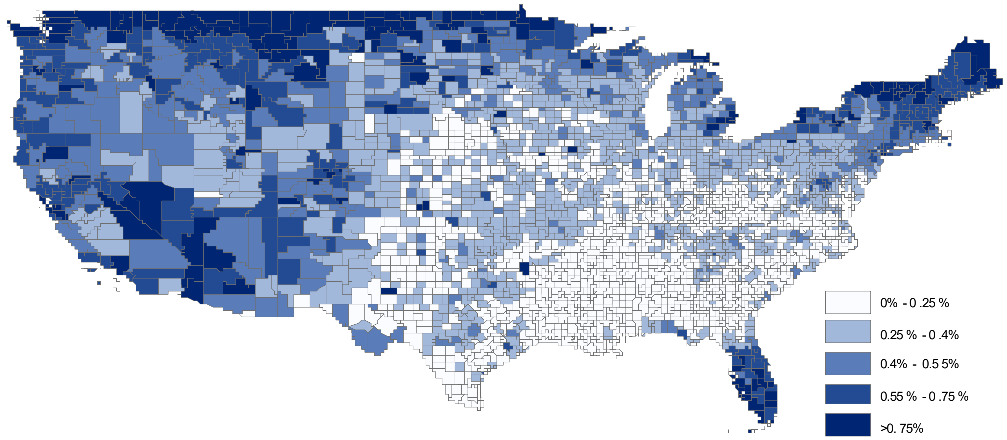
(B) Social Connectedness



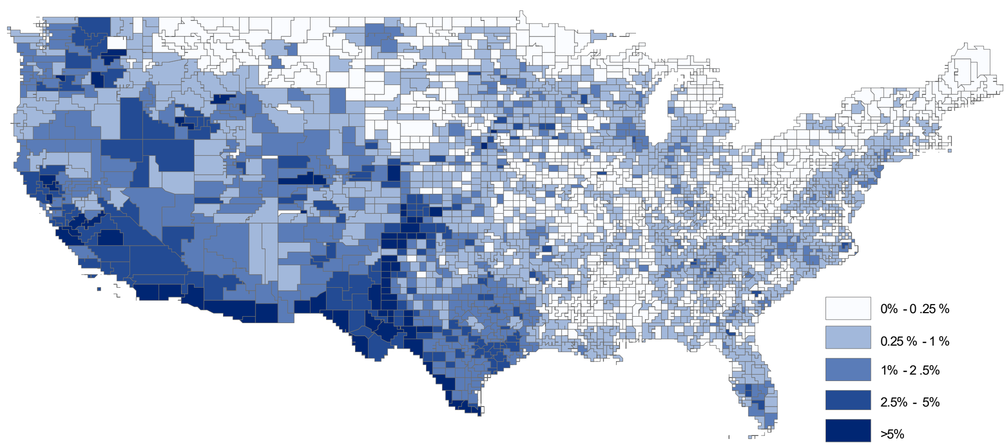
Note: Both panels show the estimated friendship elasticity of labor flows from equation A7. In Panel A, the log of the geographic distances between the states is on the horizontal axis, and in Panel B, the log of the number of friendship links is on the horizontal axis. Both panels control for state fixed effects, and Panel B also controls flexibly for the log of the geographic distance between the states.

Figure A31: International Social Connectedness

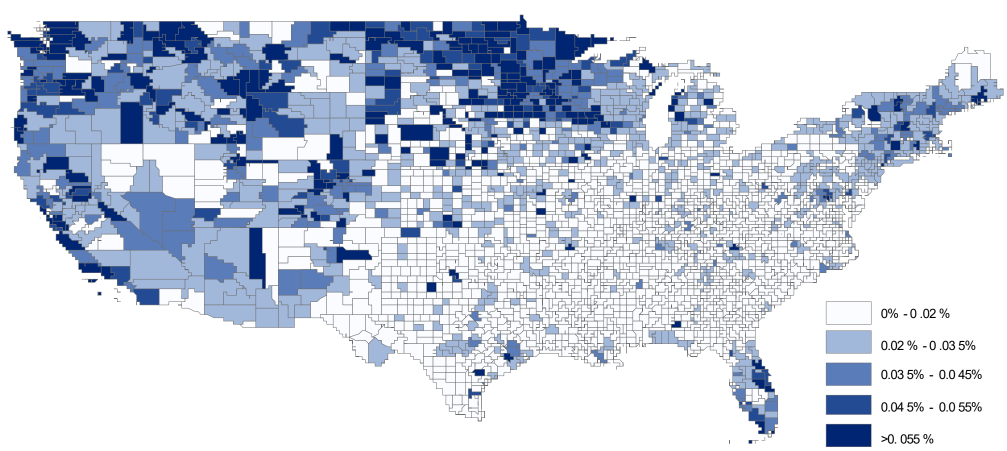
(A) Canada



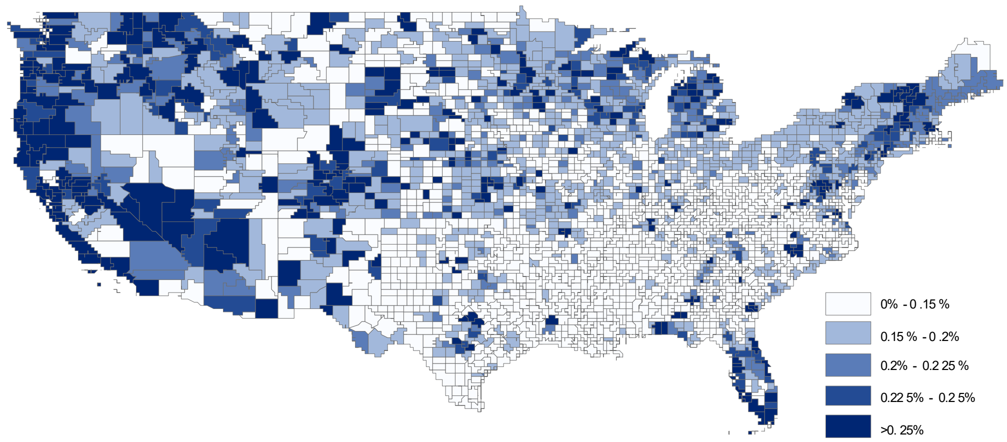
(B) Mexico



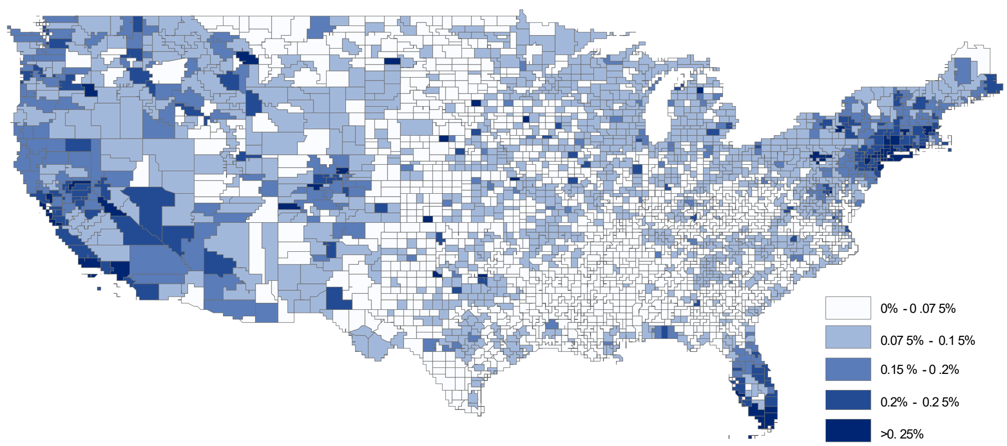
(C) Norway



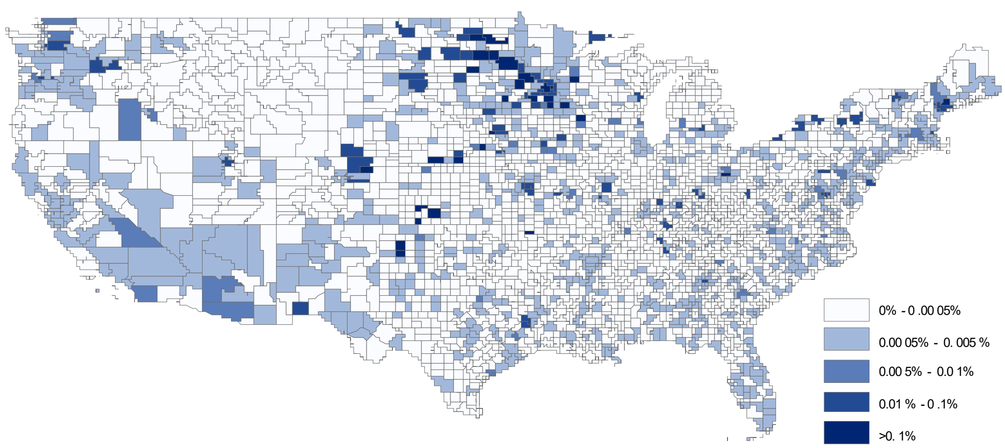
(D) Germany



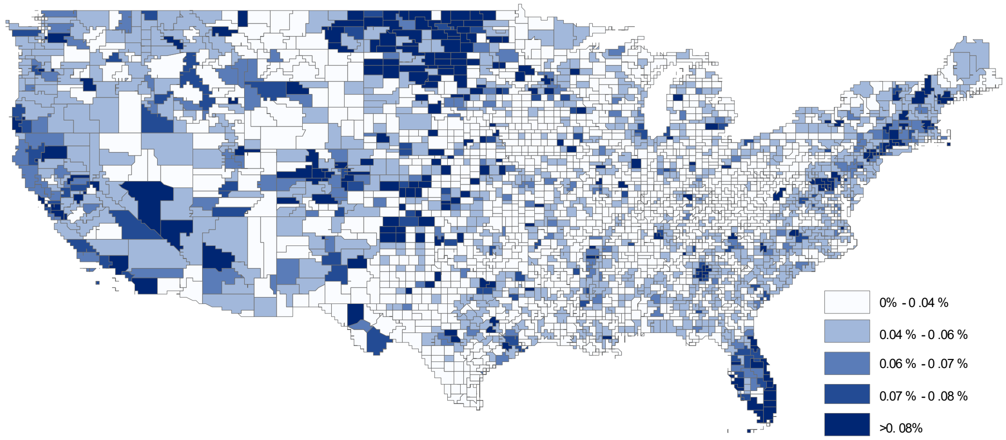
(E) Italy



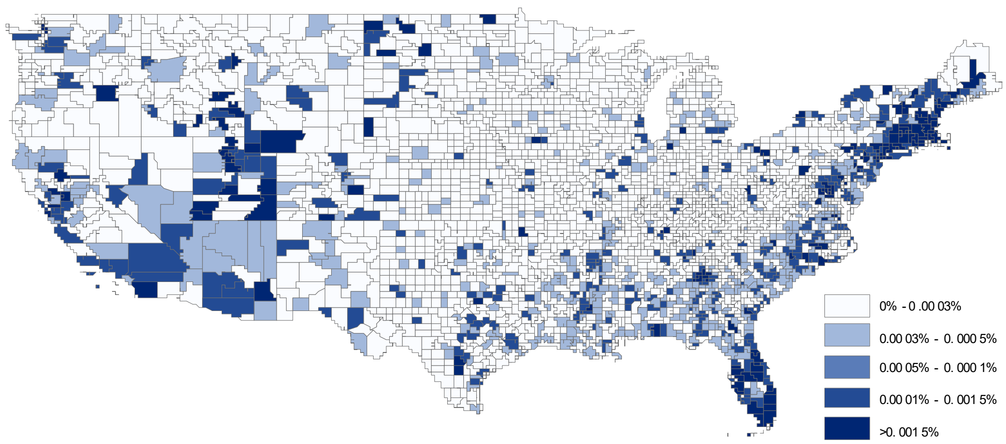
(F) Somalia



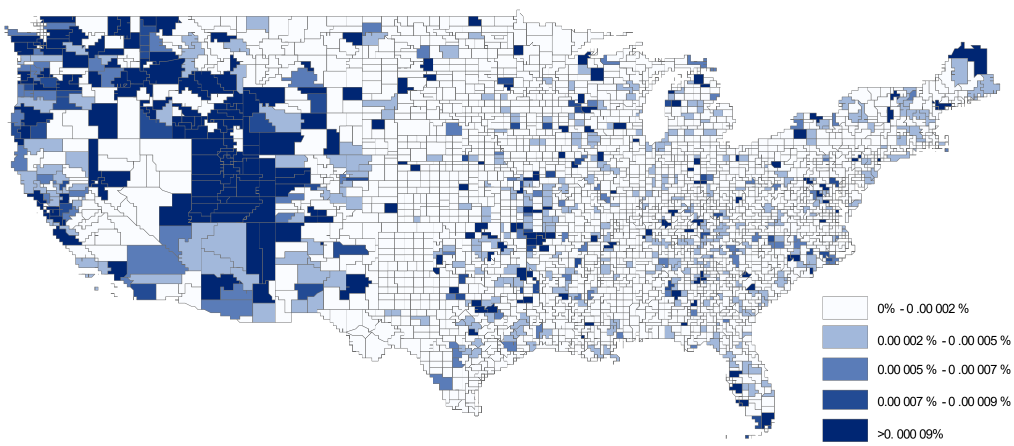
(G) South Africa



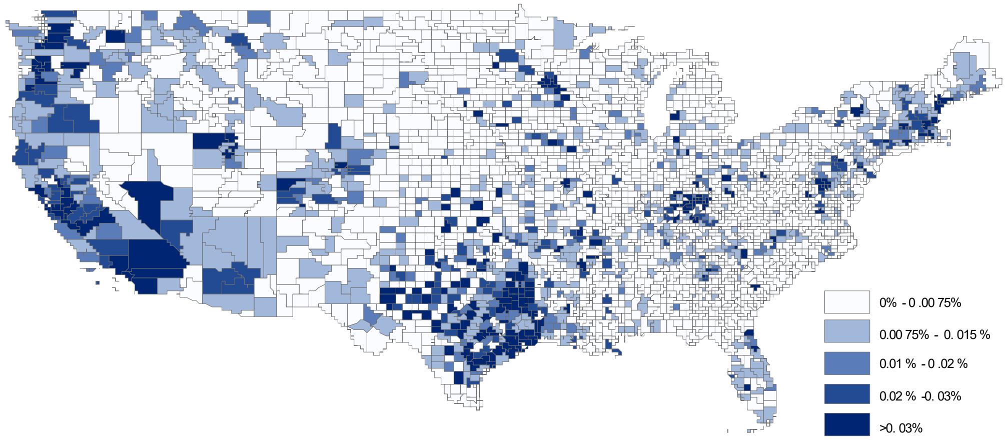
(H) Cape Verde



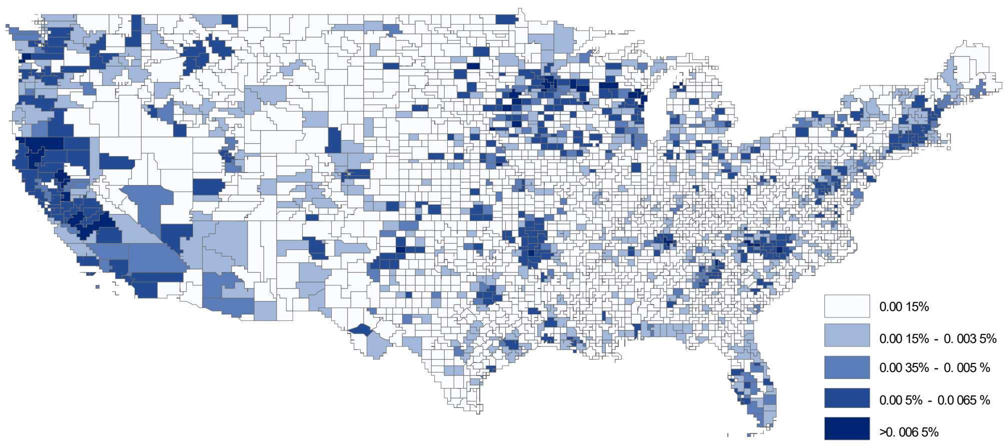
(I) Kiribati



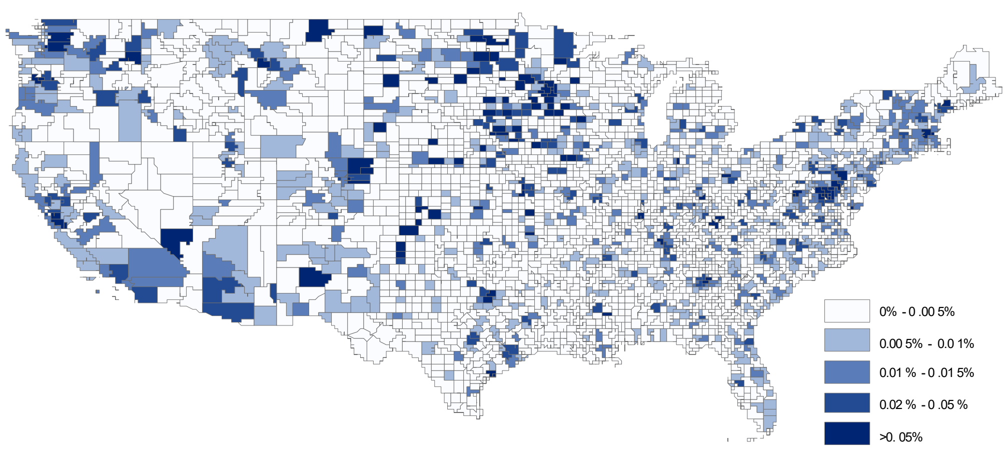
(J) Cambodia



(K) Laos



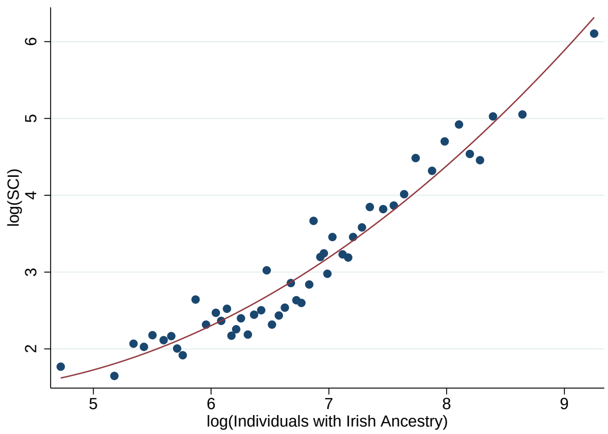
(L) Ethiopia



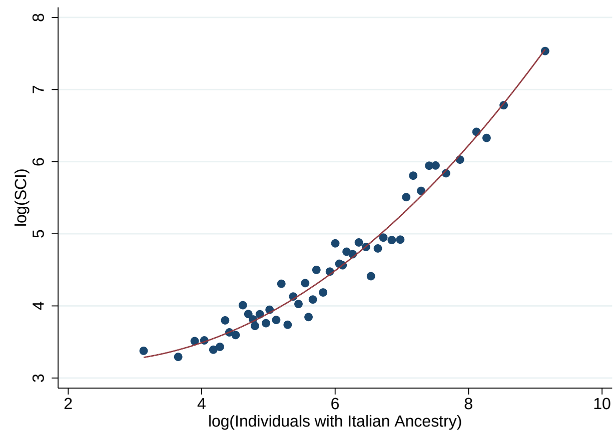
Note: Figure shows a heatmap of the share of friendship links that are to Facebook users located in Canada (Panel A), Mexico (Panel B), Norway (Panel C), Germany (Panel D), Italy (Panel E), Somalia (Panel F), South Africa (Panel G), Cape Verde (Panel H), Kiribati (Panel I), Cambodia (Panel J), Laos (Panel K), and Ethiopia (Panel L).

Figure A32: Ancestry and Social Connectedness

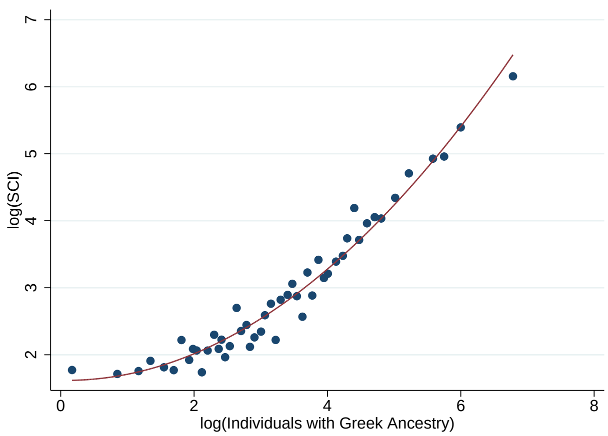
(A) Ireland - Immigration Peak in 1890 Census



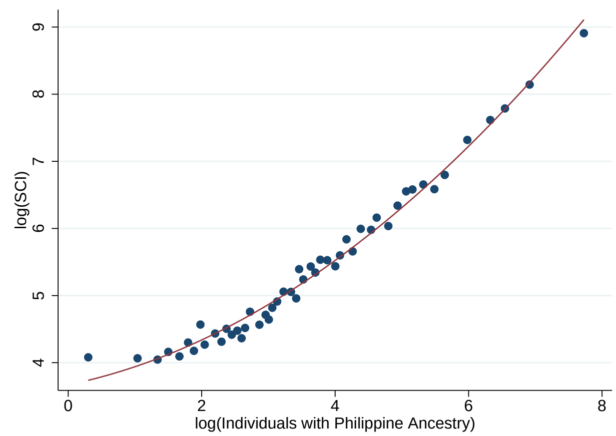
(B) Italy - Immigration Peak in 1930 Census



(C) Greece - Immigration Peak in 1980 Census

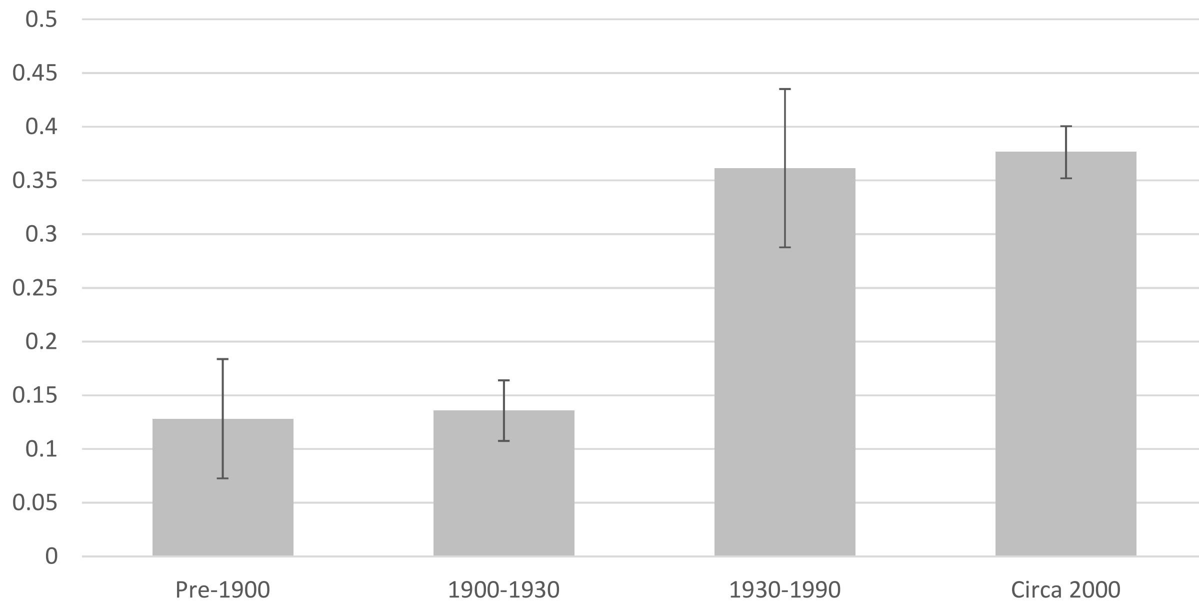


(D) Philippines - Immigration Peak in 2000 Census



Note: The figure shows binned scatter plots of the log number of residents in a county reporting a given country's ancestry on the horizontal axis and the log of the SCI between this county and the foreign country on the vertical axis. Each scatter plot controls for the log of the geographic distances between the foreign country's capital and the county.

Figure A33: Ancestry and Social Connectedness by Migration Peak



Note: Figure shows estimates of β_1 from regression A8, separately for countries grouped by the census year of peak immigration.