MEASURING SOCIAL CONNECTEDNESS*

Mike Bailey† Rachel Cao‡ Theresa Kuchler§ Johannes Stroebel¶ Arlene Wong‖

Abstract

We introduce a new measure of social connectedness between U.S. county-pairs, as well as between U.S. counties and foreign countries. Our measure, which we call the "Social Connectedness Index" (SCI), is based on the number of friendship links on Facebook, the world's largest online social networking service. Within the U.S., social connectedness is strongly decreasing in geographic distance between counties: for the population of the average county, 62.8% of friends live within 100 miles. The populations of counties with more geographically dispersed social networks are generally richer, more educated, and have a higher life expectancy. Region-pairs that are more socially connected have higher trade flows, even after controlling for geographic distance and the similarity of regions along other economic and demographic measures. Higher social connectedness is also associated with more cross-county migration and patent citations. Social connectedness between U.S. counties and foreign countries is correlated with past migration patterns, with social connectedness decaying in the time since the primary migration wave from that country. Trade with foreign countries is also strongly related to social connectedness. These results suggest that the SCI captures an important role of social networks in facilitating both economic and social interactions. Our findings also highlight the potential for the SCI to mitigate the measurement challenges that pervade empirical research on the role of social interactions across the social sciences.

JEL Codes: D8, L14, F1, O33, R23, J6

Keywords: Social Networks, Measurement, Homophily, Diffusion of Information, Patent Citations, Trade, Migration

*This version: May 9, 2017. We thank Nick Bloom, Raj Chetty, Marty Eichenbaum, Xavier Gabaix, Ed Glaeser, Erik Hurst, Seema Jayachandran, Matthew Jackson, David Laibson, Guido Lorenzoni, Brigitte Madrian, Adair Morse, and Andreas Weber, as well as numerous seminar and conference participants for useful discussions. We thank Elizabeth Casano and Patrick Farrell for outstanding research assistance. We also thank Enrico Berkes and Ruben Gaetani for sharing their patent data set. The Center for Global Economy and Business at NYU Stern for generous research support.

†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

‖Federal Reserve Bank of Minneapolis and Princeton University. Email: arlene.wong@u.northwestern.edu

Social networks shape many aspects of human life, from influencing preferences and labor market outcomes, to facilitating trade and supporting informal markets in developing economies.[1] Yet, despite a widespread recognition that interactions through social networks can have large effects on social and economic activity, the unavailability of large-scale and representative data on social connectedness between individuals or geographic regions has posed an important challenge for empirical research. Indeed, large-scale data collection on social networks has traditionally been costly and full of practical challenges, while analyses of sub-networks struggle to obtain generalizable findings.

More recently, the rise of online social networks, such as Facebook, LinkedIn, and Twitter, provides the potential to overcome some of these measurement challenges (see Bailey et al., 2016*a*). In this paper, we highlight the usefulness of such data from online social networks by introducing a new measure of social connectedness at the U.S. county level. This measure, which we call the *Social Connectedness Index* (SCI), is based on friendship links on Facebook, the world's largest online social networking service. Specifically, the SCI corresponds to the relative frequency of Facebook friendship links between every county-pair in the U.S., and between every U.S. county and every foreign country. Given Facebook's scale, with 1.8 billion active users globally and 229 million active users in the U.S. and Canada (Facebook, 2016), as well as the relative representativeness of Facebook's user body, these data provide the first comprehensive measure of friendship networks at a national level. We use these data to document important geographic patterns of social networks. We also show that the SCI data can be informative about the role of social connectedness for the large number of social and economic outcomes that can be measured at various levels of geographic aggregation, such as trade, migration, and patent citations. To facilitate further research along these dimensions, the SCI data can be made accessible to members of the broader research community.[2]

We begin by describing the construction of the SCI. We then use these new data to analyze the determinants of social connectedness between U.S. counties. We find that the intensity of friendship links is strongly declining in geographic distance, with the elasticity of the number of friendship links to geographic distance ranging from about -2.0 over distances less than 200 miles, to about -1.2 for distances larger than 200 miles. Conditional on distance, social connectedness is significantly stronger within states than across state lines. We also show that, conditional on geographic distance, the social connectedness between two counties is increasing in the similarity of these counties along important social and economic characteristics. Beyond these systematic patterns, we find that present-day friendship networks of counties are shaped by their idiosyncratic experiences, such as their exposure to large historical within-U.S. population movements: for example, Kern County, CA, has strong friendship links to the origin regions of the Dustbowl migrants that moved from Arkansas and Oklahoma to California. Cook County, IL, home to Chicago, has strong friendship links to counties along the Mississippi river, the region from which the Great Migration to northern cities originated.

We then explore the significant heterogeneity across counties in the geographic concentration of their populations' social networks. For the population-weighted average county, 62.8% of all friendship links are to individuals living within 100 miles, but this number ranges from 46.0% at the 5$^{\text{th}}$

[1] The review articles by Granovetter (2005) and Jackson (2014), as well as the handbook articles in Scott and Carrington (2011) and Bramoulle, Galeotti and Rogers (2016) provide a starting point for the interested reader.

[2] Researchers are invited to submit a one-page research proposal for working with the SCI data to sci_data@fb.com. The data will be shared for approved research projects under the terms of an NDA between Facebook and approved researchers.

percentile to 76.9% at the 95th percentile of the distribution. We analyze which characteristics of counties are correlated with this geographic concentration, and find that the populations of counties with a larger fraction of friends living more than 100 miles away are generally richer, better educated, and have higher life expectancy. These correlations suggest that controlling for the geographic concentration of social networks is likely to be important in a number of empirical analyses of economic and social activity at the county level.

In the next step, we investigate how the intensity of social connectedness between regions is related to the degree of bilateral economic and social activity. After aggregating the SCI to the state level to match available interstate trade data, we document that state-pairs with higher social connectedness see larger trade flows, even after controlling flexibly for geographic distance. This suggests that social networks help overcome some of the informational and cultural frictions that can inhibit trade. We also find that when counties are more connected, they are likely to have more cross-county patent citations. These results point to an important role of social interactions in the process of innovation, providing empirical evidence for a class of theories of economic growth that have focused on knowledge spillovers (e.g., Romer, 1986; Lucas, 1988; Aghion and Howitt, 1992). Finally, we find that more connected county-pairs see more migration and labor flows, highlighting the potential of social networks to overcome frictions involved in moving across the United States. These results complement recent research by Bailey et al. (2016*a*, 2017), who also use social network data from Facebook to document that social interactions influence people's perceptions of local housing markets as well as their real estate investment decisions and mortgage leverage choice.

We also analyze how friendship links to foreign countries correlate with both past migration patterns and present-day trade flows. We find that the social connectedness between U.S. counties and foreign countries declines with geographic distance, with similar elasticities as those estimated for within-U.S. social connectedness. We further document that past international migration patterns are important determinants of present-day social connectedness, but with elasticities that are declining in the time since the peak of the respective primary migration wave. Importantly, we also show that international trade between U.S. states and foreign countries is strongly correlated with the degree of social connectedness with those countries. This suggests that social connectedness not only helps to overcome frictions to trade within the United States, but also internationally.

Overall, the findings presented in this paper suggest that social connectedness plays a large role in explaining social and economic interactions, both within and across counties. While we focus on documenting salient patterns across a variety of settings, and do not provide full-fledged causal analyses of those patterns, our findings can guide future research on the social and economic effects of social networks. More generally, they highlight significant opportunities for using the SCI data to help alleviate the measurement challenges faced by researchers across the social sciences trying to better understand the role of social interactions.

1 Data Description

A key contribution of this paper is to construct a new measure of social connectedness for the United States. We call this measure the *Social Connectedness Index* (SCI). The SCI is constructed using aggregated and anonymized information from the universe of friendship links between all Facebook users.

Facebook was created in 2004 as an online network for college students to maintain a profile and to communicate with their friends. It has since grown to become the world's largest online social networking service, with 1.8 billion monthly active users globally, and 229 million monthly active users in the U.S. and Canada (Facebook, 2016). Duggan et al. (2015) report that as of September 2014, more than 58% of the U.S. adult population and 71% of the U.S. online population used Facebook.[3] In the U.S., Facebook mainly serves as a platform for real-world friends and acquaintances to interact online, and people usually only add connections on Facebook to individuals whom they know in the real world (Jones et al., 2013; Gilbert and Karahalios, 2009; Hampton et al., 2011).[4] Establishing a friendship link on Facebook requires the consent of both individuals, and there is an upper limit of 5,000 on the number of friends a person can add. We argue that Facebook's enormous scale, the relative reprentativeness of its user body, and the fact that individuals primarily use Facebook as a tool to interact with their real-world friends and acquaintances, account for a unique ability of the Facebook social graph to provide a large-scale representation of real-world U.S. friendship networks (see also Bailey et al., 2016*a*).

We observe an anonymized snapshot of the universe of connections between Facebook users as of April 2016. To measure the social connectedness between geographies, we map Facebook users to their respective county and country locations, and obtain the total number of friendship links between these geographies. Locations are assigned to users based on the users' regular IP address login sources. We only consider friendship links among Facebook users that have interacted with Facebook over the 30 days prior to the snapshot, and treat each friendship link identically. We then construct the SCI between all pairs of 3,136 U.S. counties, and between every U.S. county and every foreign country, as the normalized total number of friendship links for each geographic pair. In particular, the SCI is constructed to have a maximum value of 1,000,000, and relative differences in the SCI correspond to relative differences in the total number of friendship links. The highest SCI of 1,000,000 is assigned to Los Angeles County-Los Angeles County connections.

2 The Determinants of Cross-County Social Connectedness

In this section, we use the SCI data to analyze the determinants of the intensity of social connectedness between U.S. counties. We first focus on the role of geographic distance and show that there is a significant decline in the propensity of individuals to form friendship links with people living in more geographically distant counties; individuals are also more likely to form friendship links with others living in the same state. We then document that, in addition, the social connectedness between two counties is increasing in the similarity of these counties along important socioeconomic dimensions. Finally, we explore which groups of counties exhibit the strongest community ties.

We first analyze the role of geographic distance in shaping social connectedness in the United

[3]Duggan et al. (2015) also report that among online U.S. adults, Facebook usage rates are relatively constant across income groups, education groups, and racial groups. Usage rates among online U.S. adults are declining in age, from 87% of 18-to-29-year-olds to 56% of above-65-year-olds.

[4]The survey by Duggan et al. (2015) asked individuals to characterize their friendship network: 93% of Facebook users said they are Facebook friends with family members other than parents or children; 91% said they are Facebook friends with current friends; 87% said they are connected to friends from the past, such as high school or college classmates. 58% said they are connected to work colleagues; 45% said they are Facebook friends with their parents; 43% said they are friends with their children on Facebook; 36% said they are Facebook friends with their neighbors. Only 39% reported to have a Facebook connection to someone they never met in person.

States. The effects of geographic proximity on friendship formation and social interactions have been studied in a number of important papers, including Zipf (1949), Holahan et al. (1978), Verbrugge (1983), and Marmaros and Sacerdote (2006). The SCI allows us to re-examine this relationship using detailed and large-scale information on social connectedness at a national level.

As a motivating example, the maps in Figure 1 show the intensity of friendship links of San Francisco County, CA (Panels A and C), and of Kern County, CA (Panels B and D), with all other counties in the continental United States. In Panels A and B, we plot the share of friendship links from home county i to each other county j, where $i \in \{\text{San Francisco, Kern}\}$. This measure is constructed as:

$$ShareFriends_{i,j} = \frac{SCI_{i,j}}{\sum_j SCI_{i,j}}. \tag{1}$$

This measure will, by construction, be larger for counties j with a larger population. We therefore also create a second measure that is independent of the size of the target county. We construct this measure as the SCI between county i and j, divided by the product of the number of Facebook users in counties i and j:[5]

$$RelativeProbFriendship_{i,j} = \frac{SCI_{i,j}}{FB\_Users_i \times FB\_Users_j}. \tag{2}$$

This measure captures the relative probability that a given Facebook user in county i is connected to a given Facebook user in county j. In Panels C and D, we plot scaled versions of $RelativeProbFriendship_{i,j}$. Due to the scaling of the SCI, only relative magnitudes of this variable can be interpreted: if it is twice as large, a given Facebook user in county i is twice as likely to be connected with a given Facebook user in county j.

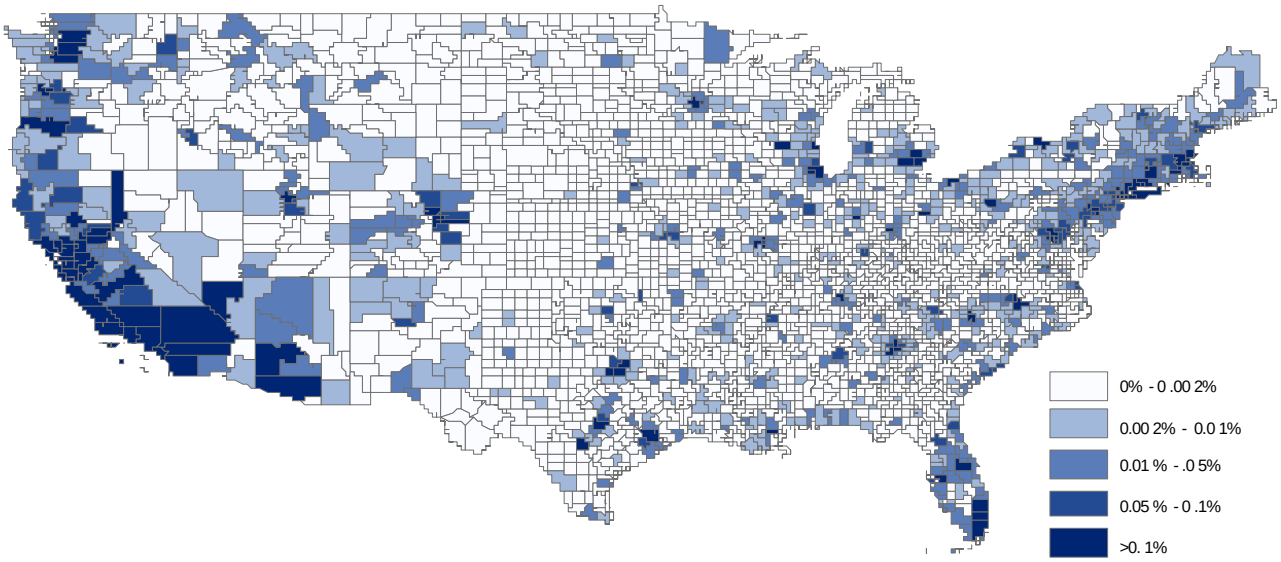
For both San Francisco County and Kern County, a significant proportion of friendship links is to geographically close counties. In addition, both counties have friendship links across the West Coast of the United States. However, there are also noticeable differences in the patterns of social connectedness: while the population of San Francisco County has significant social connections to counties located in the north-eastern United States, the population of Kern County has far fewer of these friendship links. Instead, Kern County's friendship network is very concentrated in the West Coast and Mountain States, with the exception of a pocket of strong connections to individuals living in Oklahoma and Arkansas. These connections are likely related to past migration patterns: Kern County was a major destination for migrants fleeing the Dust Bowl in the 1930s, and half of the residents of the San Joaquin Valley (within which Kern County lies) have ancestors who migrated from affected regions (see News OK, 2015, for more information). There are also disproportionately many friendship links between Kern County and the oil-producing regions of North Dakota, perhaps not surprising given that Kern County produces more oil than any other county in the United States (see Los Angeles Times, 2016, for more information). Overall, we find that the friendship networks of the Kern County population are much more geographically concentrated than those of the San Francisco County population: Kern County has 57% of friends living within 50 miles (and 75% within 200 miles), relative to 27% (48%) for San Francisco County.[6]

[5]The public-release version of the data does not contain information on the number of Facebook users per county. However, very similar results are obtained when dividing the SCI by the product of the county-level populations.

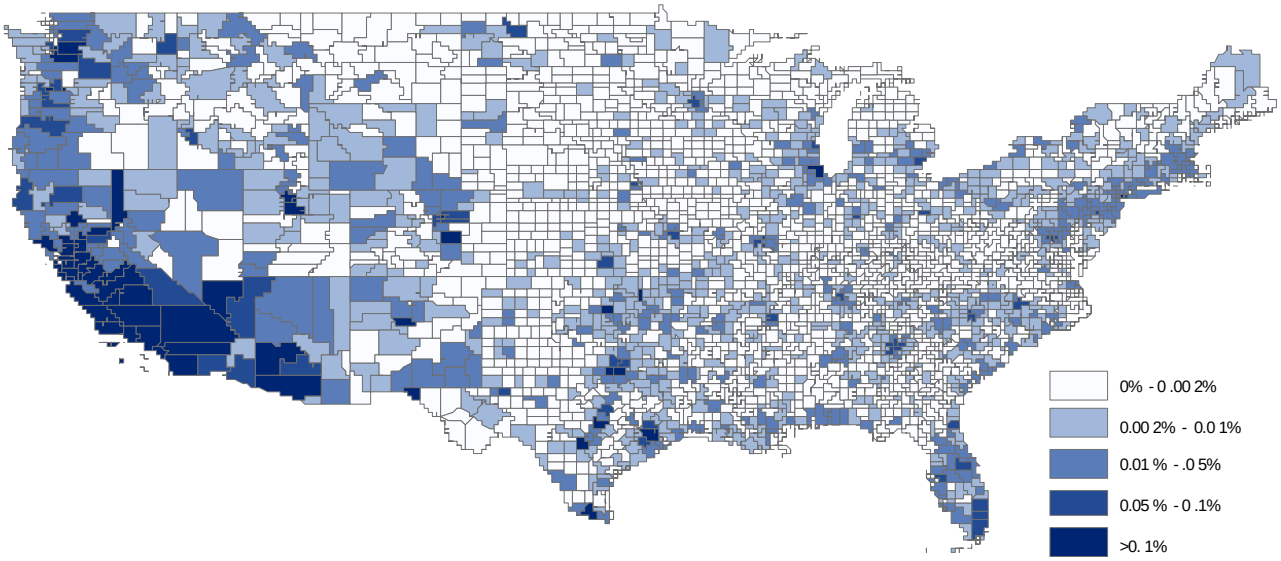
[6]Similar divergence in the geographic dispersion of friendship networks can be seen in Figure A1, which maps the

Figure 1: County-Level Friendship Maps

(A) San Francisco County, CA - Share of Friendship Links ($ShareFriends_{i,j}$)



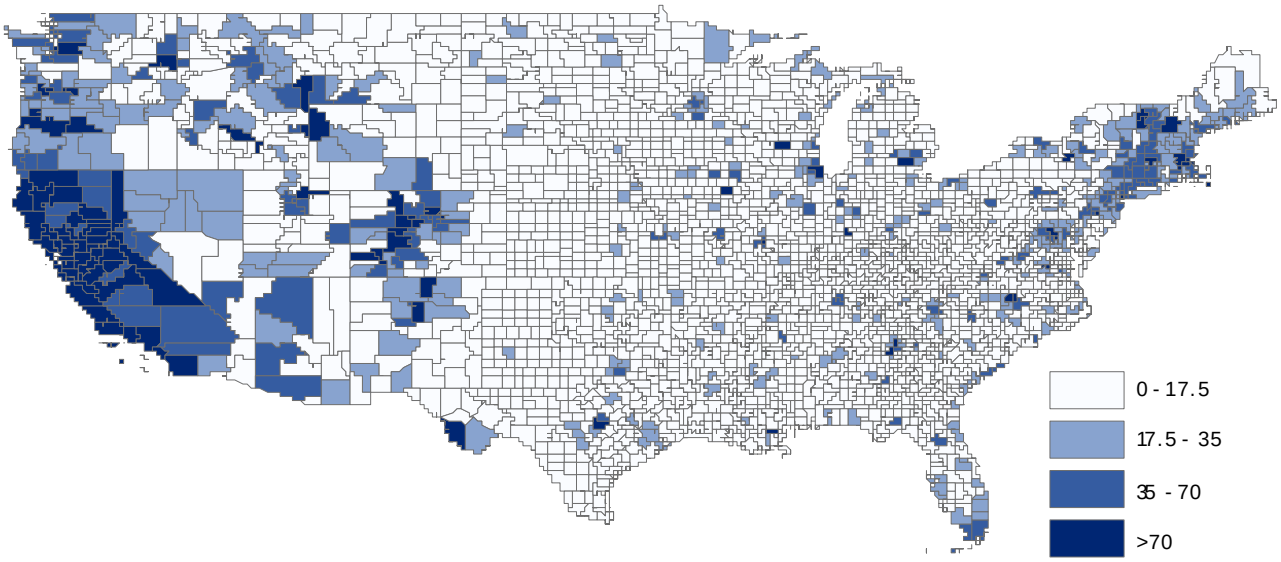
(B) Kern County, CA - Share of Friendship Links ($ShareFriends_{i,j}$)



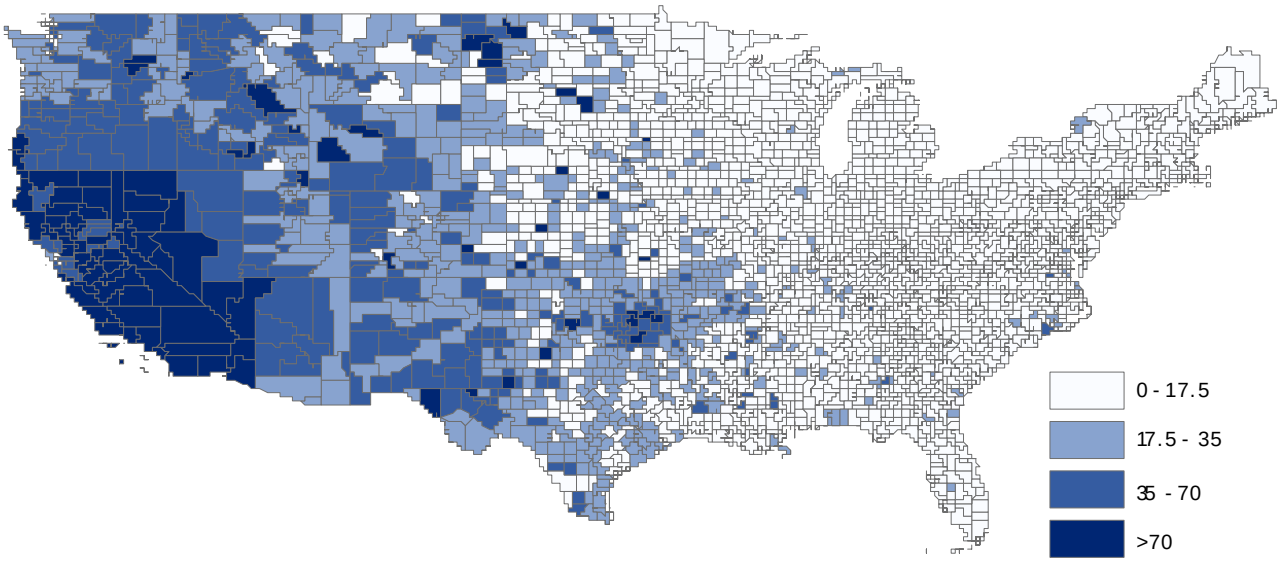
Note: Figure shows the share of friendship links of San Francisco County, CA (Panel A) and Kern County, CA (Panel B) to all other counties in the continental United States, constructed as in equation 1. Darker colors correspond to counties in which the home county i's Facebook users have a larger share of friends.

Figure 1: County-Level Friendship Maps

(C) Relative Probability of Friendship Link to San Francisco County, CA ($RelativeProbFriendship_{i,j}$)



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



Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to San Francisco County, CA (Panel C) and Kern County, CA (Panel D). It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (San Francisco or Kern) and county j.

Table 1: 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 U.S. population living within a certain distance. Counties are weighted by their population.

Table 1 shows that the geographic concentration of the friendship network of Kern County is relatively representative of the U.S. average, while San Francisco County's friendship network is extremely geographically dispersed. For the average (population-weighted) U.S. county, 55.4% of friends live within 50 miles, with a 10-90 percentile range of 42.5% to 67.4%. For the average county, over 70% of friends live within 200 miles, with a 10-90 percentile range of 57.1% to 81.2%. This is despite the fact that, for the average county, only 1.3% and 6.6% of the U.S. population live within 50 miles and 200 miles, respectively.[7]

Figure 2 illustrates the strength of friendship links between 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 2, 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 U.S. 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.

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

friendship networks for Manhattan and the Bronx.

[7] Table 1 reflects the concentrations of friendship networks across all 50 states. Table A1 contains the concentrations for only the contiguous states (excluding Alaska and Hawaii), which are roughly the same.

Figure 2: State-State Adjacency Matrix of Friend Links

Note: Figure shows an adjacency matrix of the probability of social connections, constructed as in equation 2, 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).

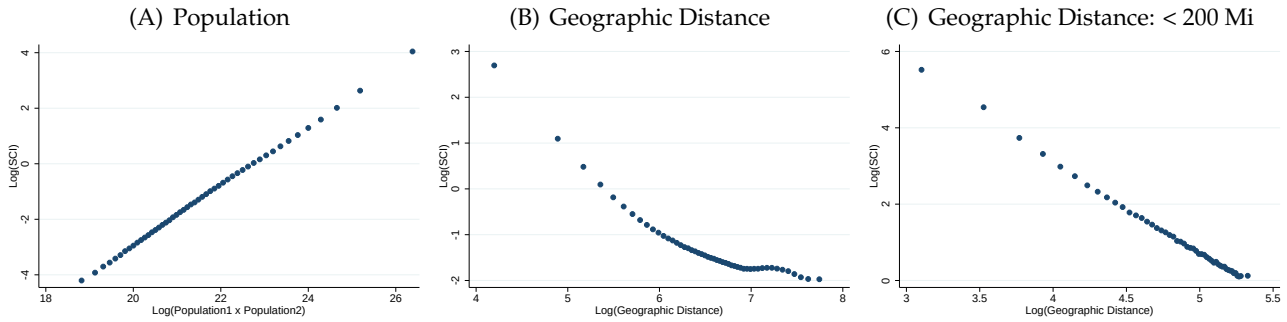
in the United Kingdom (Lambiotte et al., 2008), estimates of -1 among bloggers (Liben-Nowell et al., 2005), and estimates of -0.5 in location-based online social networks such as Brightkite, Foursquare, and Gowalla (Scellato et al., 2011).

To analyze whether a similar relationship holds for friendships at the county level, we need to control for the total populations in each of the two counties, since counties with larger populations are more likely to have more friendship links. To see this, Panel A of Figure 3 presents a binned scatter plot at the county-pair level.[8] On the vertical axis we plot the log of the number of friendship links (i.e., the log of the SCI) between the counties, and on the horizontal axis we plot the log of the product of the two counties' populations. There is a strong, linear relationship: all else equal, county-pairs with larger populations across the two counties have more friendship links between them.

[8]We drop all county-pairs where either county has a population of fewer than 10,000 people in this and all subsequent analyses in this section.

In Panel B of Figure 3, we plot a similar binned scatter plot, this time 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.[9] Conditional on the population, two counties have more friendship links when they are closer geographically. However, the relationship appears to be non-linear, with a more negative elasticity at shorter geographic distances between the two counties. In Panel C of Figure 3, we again plot the relationship between geographic distance and friendship links, now restricting the sample to county-pairs that are less than 200 miles apart. In this range of the distance, which includes about 70% of all friendship links, the elasticity of social connectedness to geographic distance is nearly constant.

Figure 3: County-Level Social Connectedness



Note: Figure shows binned scatter plots with county-pairs as the unit of observation. In Panel A, the log of the product of the county populations is on the horizontal axis, and the log of the SCI is on the vertical axis. Panel B shows a conditional binned scatter plot, where we flexibly condition on the log of the product of the populations in the two counties; 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. Panel C shows a subset of Panel B focused on county-pairs that are less than 200 miles apart.

To obtain magnitudes for the associated elasticities, we estimate regression 3. The unit of observation is a county-pair. The dependent variable, $\log(f_{ij})$, denotes the log of the number of friendship links between counties i and j (i.e., the log of the SCI); $\log(pop_i \times pop_j)$ 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.

$$\log(f_{ij}) = \beta_0 + \beta_1 \log(pop_i \times pop_j) + \beta_2 \log(d_{ij}) + \epsilon_{ij} \tag{3}$$

Column 1 of Table 2 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. In column 3, we include fixed effects for counties i and j. This absorbs $\log(pop_i \times pop_j)$ as a regressor, and controls for any other characteristics that vary at the county level. In this specification, the estimated elasticity of social connectedness to geographic

[9] We only focus on the continental United States. We condition on the log of the product of the counties' populations by including 50 dummy variables for equal-sized percentiles of the distribution.

distance is about -1.48. This estimate suggests that a 10% increase in the distance between two counties is associated with a 14.8% decline in the number of friendship links between those counties.

In column 4, we include an additional control indicating when both counties are within the same state. The social connectedness of a county is often strongest with other counties within the same state, even compared to nearby counties in other states. Indeed, state borders can regularly be identified when mapping the social connectedness of a county (see Appendix Figure A8). Why state borders play such an important role in determining social connectendess, and the extent to which this is driven by institutional, social, or economic factors, is an interesting avenue for future research.

In columns 5 and 6, we restrict the sample to county-pairs that are more and less than 200 miles apart, respectively. In the sample of county-pairs that are less than 200 miles apart, the same sample as in Panel C of Figure 3, the estimated elasticity between geographic distance and friendship links is -1.98, suggesting that a 10% increase in the geographic distance between two counties would lead to a roughly 20% reduction in the number of friendship links between them. In the sample of county-pairs that are more than 200 miles apart, the magnitude of the elasticity falls by nearly half to -1.16. These findings confirm that while social connectedness is declining in geographic distance, the elasticity of this relationship is less negative as we include county-pairs that are progressively further apart. This suggests that in the theoretical modeling of friendship links, the appropriate elasticity depends on the geographic distances studied.

A substantial literature has documented that individuals are more likely to be associated with other individuals of similar characteristics. Following Lazarsfeld and Merton (1954), this empirical regularity is referred to as "homophily." Homophily has been documented for a large number of individual characteristics, including racial identity, gender, age, religion, and education, as well as intangible aspects such as attitudes and beliefs (see McPherson, Smith-Lovin and Cook, 2001, for a comprehensive review of the literature). The presence of such homophily can have important effects. For example, it can affect preferences (Rosenblat and Mobius, 2004), and it can slow down the speed of learning and reaching agreement on issues of broad interest (Golub and Jackson, 2012).

The previous findings suggest that geographic proximity is an important determinant of friendship links between two counties. This can be interpreted as a first dimension of county-level homophily: people are more likely to be friends with others in counties that are similar in terms of geographic location. We next analyze whether we can detect additional dimensions of county-level homophily. In particular, we estimate the degree to which the social connectedness of two counties depends on the similarity of these counties along socioeconomic dimensions such as income and education. In doing so, we are looking for forces that explain social connectedness between counties over and above what would be predicted purely by the geographic distance between these counties.

To do this, we expand regression 3 to also include measures of the differences between county-pairs along socioeconomic dimensions. As before, we include fixed effects for counties i and j, and a dummy variable indicating when both counties are in the same state. Column 7 confirms that the number of friendship links is indeed correlated with the degree of similarity of counties along a number of the socioeconomic measures.[10] Importantly, the estimated elasticity of friendship links with

[10]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

Table 2: Determinants of Social Connectedness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
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***	-1.094***	-1.951***
		(0.064)	(0.065)	(0.061)	(0.059)	(0.043)	(0.055)	(0.051)	(0.047)
Same State				1.496***	1.271***	1.216***	1.496***	1.283***	1.210***
				(0.087)	(0.083)	(0.044)	(0.085)	(0.086)	(0.044)
Δ Income (k$)							-0.006***	-0.007***	0.003***
							(0.001)	(0.001)	(0.001)
Δ Share Pop White (%)							-0.012***	-0.012***	-0.008***
							(0.001)	(0.001)	(0.001)
Δ Share Pop No High School (%)							-0.012***	-0.011***	-0.011***
							(0.002)	(0.002)	(0.003)
Δ 2008 Obama Vote Share (%)							-0.006***	-0.006***	-0.006***
							(0.001)	(0.001)	(0.001)
Δ Share Pop Religious (%)							-0.002***	-0.002***	-0.002**
							(0.001)	(0.001)	(0.001)
County Fixed Effects	N	N	Y	Y	Y	Y	Y	Y	Y
Sample					>200 miles	<200 miles		>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	2,775,244	186,669
R2	0.682	0.813	0.907	0.916	0.916	0.941	0.922	0.922	0.943

Note: Table shows results from regression 3. 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. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

respect to geographic distance remains relatively unaffected. 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 deviations) 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 4 and 7 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.[11]

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.

[11] Figure A7 shows binned scatter plots at the county-pair level that portray the relationship between differences across

In columns 8 to 9 we explore how the relationships between socioeconomic differences and social connectedness vary with the geographic distance between the counties, again by splitting the sample into county-pairs that are more and less than 200 miles apart. As before, the elasticity of social connectedness to geographic distance is more negative when focusing on county-pairs that are closer together. The relationship between income and social connectedness actually flips sign, while the elasticity of social connectedness to the other county-level differences does not appear to be significantly different across shorter or longer geographic distances.

In this section, we have highlighted a number of important forces that correlate with the social connectedness across counties: social connectedness is decreasing in geographic distance, and declines notably across state borders. Counties are also more likely to be socially connected if they are more similar on a number of important socioeconomic dimensions. In Appendix A we explore a number of other, more idiosyncratic determinants of friendship links across counties. We document that the strength of social connections may be affected by physical obstacles such as large rivers and mountain ranges. We highlight that counties with military bases exhibit strong connections across the entirety of the United States, as do counties in North Dakota that have seen a recent shale oil boom and an associated significant in-migration. We show that counties with Native American reservations are strongly connected to each other. Similarly, areas with ski resorts in the Rocky Mountains and New England are strongly connected to each other. We also find that counties in Florida with significant retiree populations are strongly connected to the Rustbelt and the Northeast. In addition, large cities in the Midwestern United States with significant African American populations, such as Milwaukee and Chicago, have strong links to the South around Mississippi and Alabama, consistent with friendship links persisting following the Great Migration of southern African Americans to northern cities.

2.1 Connected Communities Within the United States

Table 2 highlights that social connectedness drops off strongly at state borders. A related question is how closely existing state borders resemble the borders that would form if we grouped together U.S. counties to create communities with the aim of maximizing within-community social connectedness.[12] There are a number of possible algorithms to facilitate such a grouping of counties. In our application, we use hierarchical agglomerative linkage clustering. Conceptually, this algorithm starts by considering each of the N counties in the U.S. 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 2: 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.

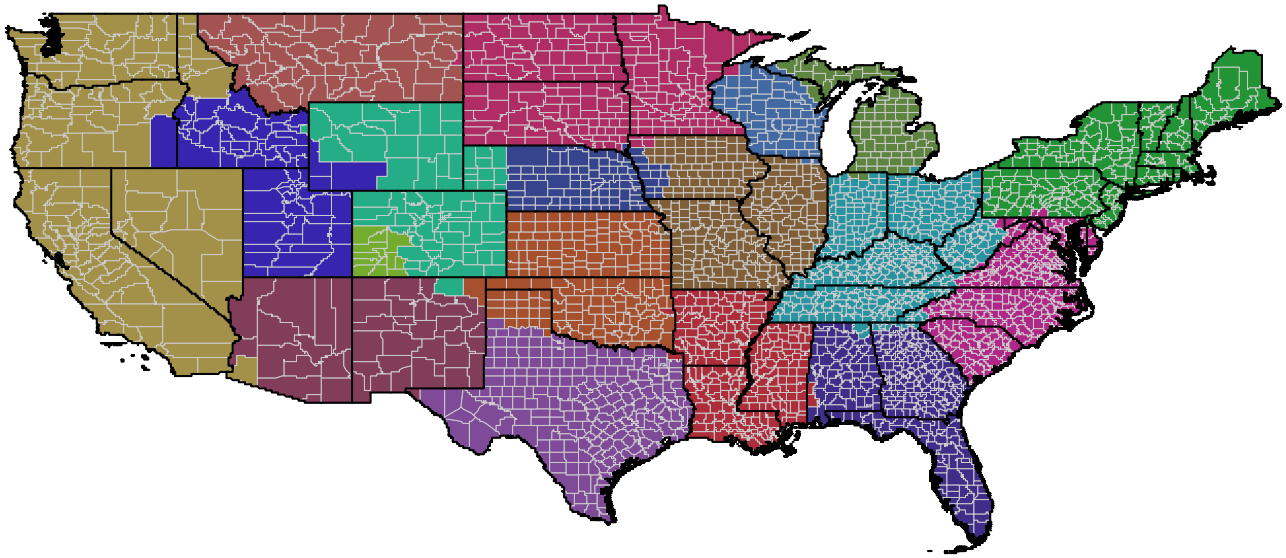
Panel A of Figure 4 shows the result when we use this algorithm to group the United States into 20 distinct communities. All resulting communities are spatially contiguous, despite this not being

the two counties along a number of important outcome variables, and their social connectedness. Most of the relationships are relatively linear.

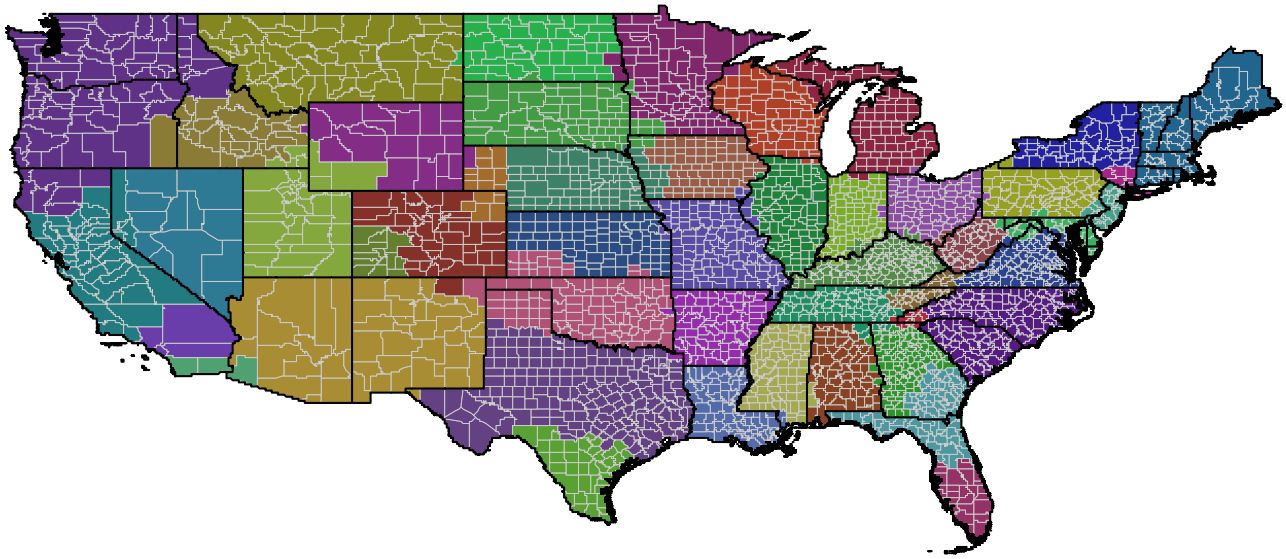
[12] In recent related work, Calabrese et al. (2011) document spatial community structures based on U.S. cellphone data (see also Ratti et al., 2010; Blondel et al., 2010).

Figure 4: Connected Communities within the United States

(A) 20 Distinct Units



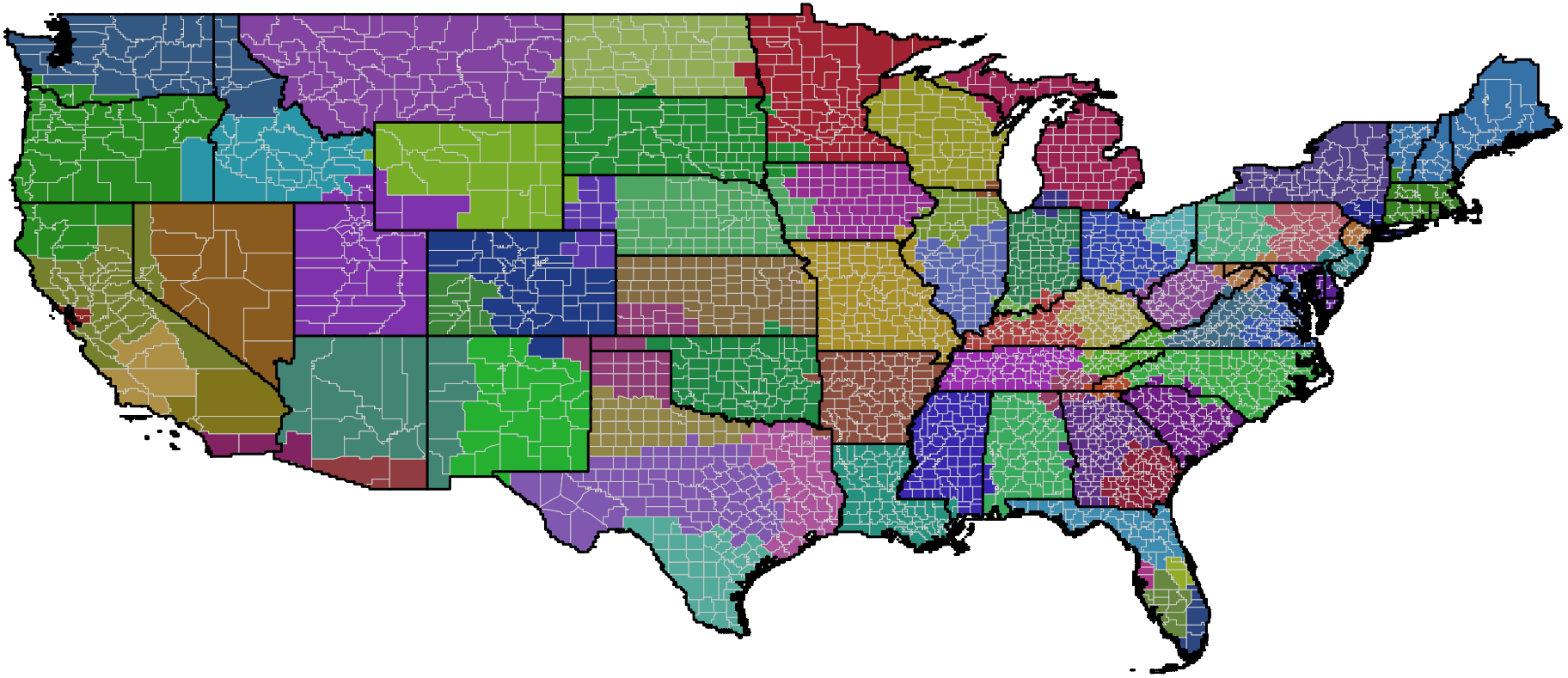
(B) 50 Distinct Units



Note: Figure shows U.S. counties grouped together when we use hierarchical agglomerative linkage clustering to create 20 (Panel A) and 50 (Panel B) distinct groups of counties. The algorithm assigns both Hawaii and Alaska, not pictured, to two distinct clusters including only the respective state in Panels A, B, and C.

Figure 4: Connected Communities within the United States

(C) 75 Distinct Units



Note: Figure shows U.S. counties grouped together when we use hierarchical agglomerative linkage clustering to create 75 distinct groups of counties (Panel C). The algorithm assigns both Hawaii and Alaska, not pictured, to two distinct clusters including only the respective state in Panels A, B, and C.

a constraint enforced by the clustering algorithm; this is a result of the strong dependence of social connectedness on geographic distance. In addition, and consistent with finding social connectedness to decline at state borders, many of the community borders line up with state borders. All of the West Coast States together with Nevada form one community. Similarly, all counties in states between New England and Pennsylvania are grouped into the same community. Other groups of states are Florida, Georgia and Alabama, as well as Louisiana, Arkansas and Mississippi. Tennessee, Kentucky, and West Virginia are grouped with Ohio and Indiana; Illinois is grouped with Iowa and Missouri. Michigan, Wisconsin, and Montana are each their own community, while northern Texas is grouped together with Oklahoma and Kansas. A small separate community is formed by Colorado's Western Slope region, a region the Denver Post (2010) has referred to as the "Other Colorado," an apt description of its appearance in the Figure.[13]

In Panel B of Figure 4, we group the United States into 50 distinct communities. Many multi-state groups from Panel A 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.

In Panel C of Figure 4 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 Panel B of Figure 4. 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.

Overall, this section highlights the ability of the SCI data to describe important patterns of within-U.S. social connectedness. It also shows how researchers can use these new data to better understand the forces driving that connectedness.

[13]The full quote reads: "A trip to the Western Slope is a visit to a different world. People across the divide call their rural home 'The Other Colorado' and are quick to tell you why they live here and why they came - less traffic, more leisurely lifestyle, milder weather, the intimacy of a small-town community."

3 Concentration of Social Networks and County Characteristics

The previous section documents that, on average, the number of friendship links between two counties is declining in the geographic distance between the counties. However, Table 1 reveals significant heterogeneity in how geographically concentrated the friendship networks of various counties are: the 5-95 percentile range across population-weighted counties in the share of friends living within 100 miles is 46.0% to 76.9%. Existing theoretical work suggests that the diversity of social networks is an important determinant of economic development, and that tightly clustered social ties can limit access to a broad range of social and economic opportunities (e.g., Granovetter, 1973; Page, 2008). Yet, empirical studies of the relationship between the structure of social networks and economic outcomes of communities are rare. The exception is Eagle, Macy and Claxton (2010), who use U.K. cell phone data to document that the diversity of individuals' social networks is correlated with regional economic well-being. In this section, we provide evidence that the geographic dispersion of friendship links in the U.S. is highly correlated with social and economic outcomes at the county level, such as average income, educational attainment, and social mobility.

We measure the concentration of social networks in two ways. The geographic concentration of social networks is measured as the share of friends that live within 100 miles of a county, and the density of social networks is measured as the share of friends among the nearest 50 million people in and surrounding a county.[14] The maps in Figure 5 shows the two measures of the concentration of different counties' social networks, with darker areas corresponding to more concentrated networks. There are notable differences between Panel A, which shows the geographic concentration of social networks, and Panel B, which shows the density of social networks. Overall, friendship networks in the South, the Midwest, and Appalachia appear the most geographically concentrated. Counties in the Rocky Mountains, a less-densely populated area of the U.S., have the smallest share of friends living within 100 miles. Among the western United States, Utah and inland California have the most geographically concentrated friendship networks. In Panel B, 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. Differences in the two measures of concentration are the result of variation in population density across the United States.

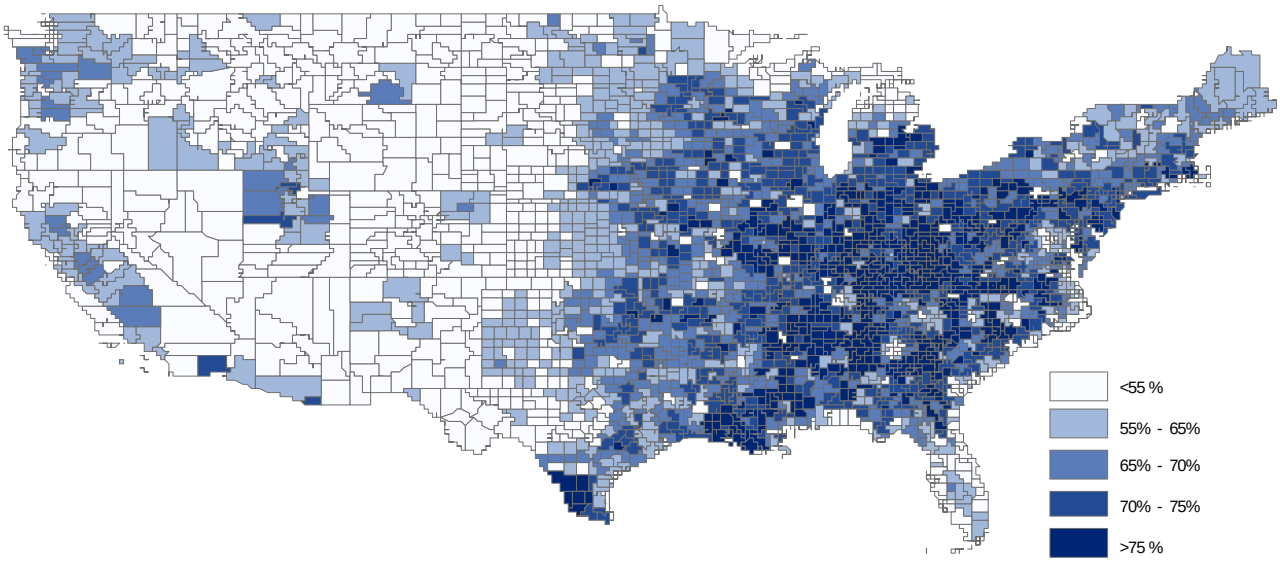
What are the effects of differentially structured social networks on county-level outcomes? As a first step toward answering this question, we next correlate our measures of the concentration of friendship links with county-level social and demographic characteristics. Importantly, these correlations cannot by themselves be interpreted as causal. Our goal here, as in the rest of the paper, is to document a number of stylized facts that can guide future research that investigates causal effects of social network structure on socioeconomic outcomes. We also aim to demonstrate the power of the SCI data to overcome the measurement challenges to such research.

Figure 6 presents county-level binned scatter plots of the share of friends living within 100 miles by demographic characteristics. Panel A shows that counties with higher average income have more dispersed friendship networks. The relationship is not linear: mean household incomes are roughly

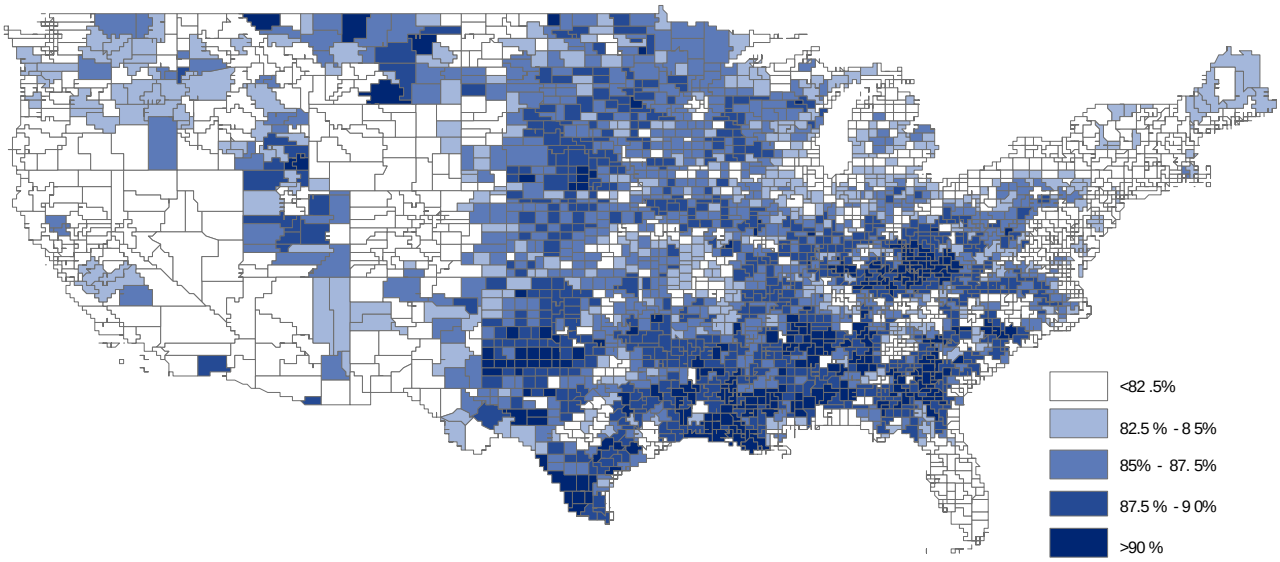
[14] We find similar results when looking at the share of friends within 50, 100, 300, and 500 miles, the share of friends among the nearest 10, 25, and 100 million people, and when we measure concentration using a county-level Herfindahl index of friendship links.

Figure 5: Concentration of Social Networks

(A) Share of Friends Living within 100 Miles

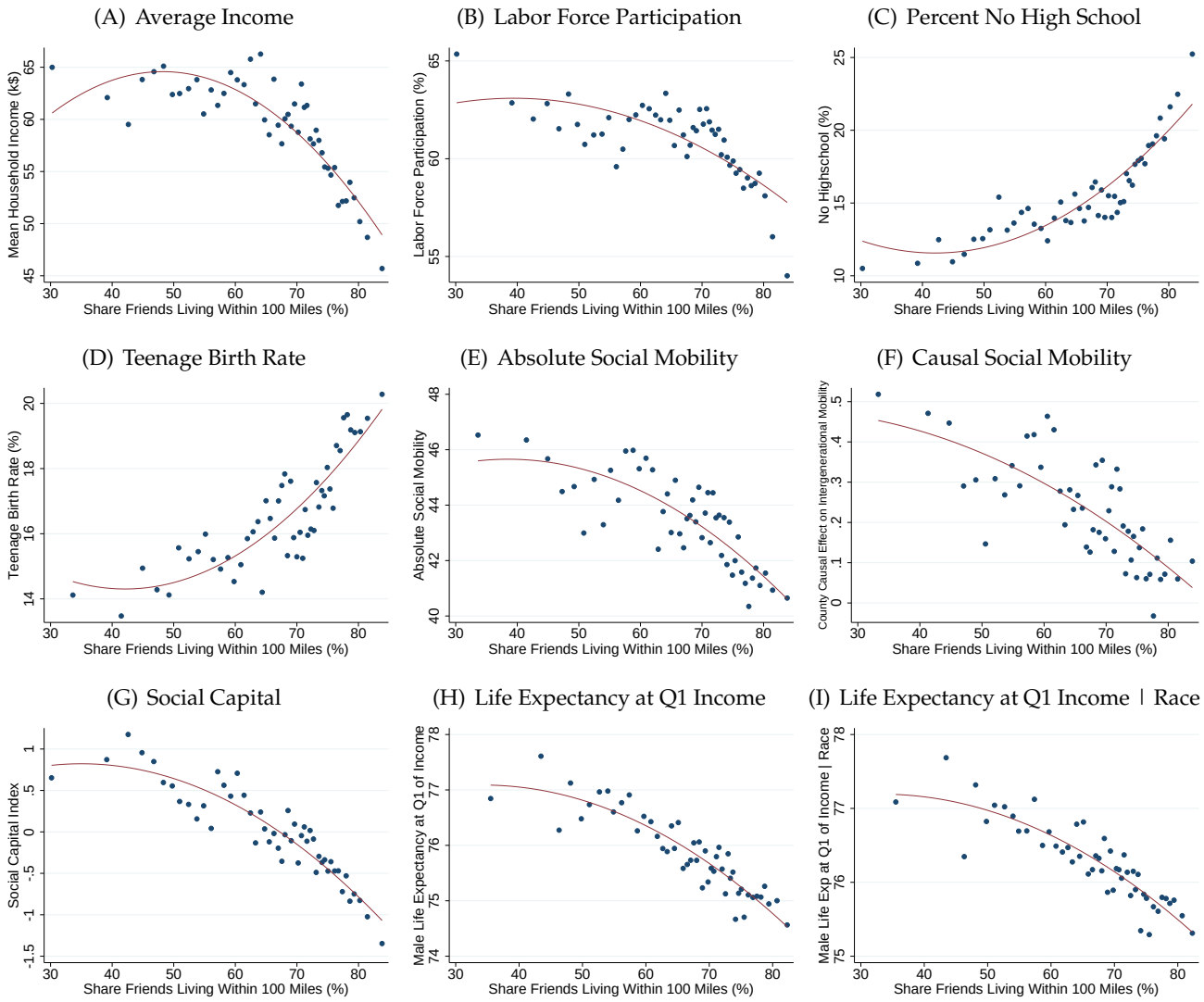


(B) Share of Friends Among Nearest 50 Million People



Note: Panel A shows a map at the county level of the share of all U.S. friends that live within 100 miles. Panel B shows a map at the county level of the share of all U.S. friends that are among the nearest 50 million people.

Figure 6: 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.

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 Figure 6 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. When comparing across counties, 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 Figure 6, 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 Figure 6 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.[15]

[15] Appendix Figures A2 and A3 show similar relationships as in Figure 6, but 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 relationships between the geographic dispersion of friendship networks and socioeconomic outcomes also hold when including state and commuting zone fixed effects. Interestingly, Panels E and F of Appendix Figures A2 and A3

Appendix Figure A4 shows county-level binned scatter plots for the same demographic variables as Figure 6 using the share of friends within the nearest 50 million people rather than the share of friends within 100 miles.[16] For some of the demographic characteristics, the geographic concentration of social networks shown in Figure 6 has greater predictive power while for others the density of social networks as shown in Appendix Figure A4 is a stronger predictor. The R^2 of the quadratic regressions that underlie each of the Panels in Figure 6 [Appendix Figure A4] 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.

The previous analysis 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. In Appendix, we present results of a multivariate regression of our measures of the concentration of social networks on our county-level outcome measures. We find that most of the relationships persist in a multivariate analysis.

This section has documented a strong relationship between the geographic dispersion of county-level friendship networks and county-level economic, social, and health outcomes. We hope that future work will investigate the extent and direction of causality for these relationships. In addition, the newly available SCI data should allow researchers to measure the diversity of county-level social networks not just along geographic dimensions, but also along cultural, political, and socioeconomic dimensions. More generally, the strong correlation between social connectedness and socioeconomic outcomes suggests that controlling for the geographic concentration of social networks is important to minimize omitted variables bias across a number of research agendas that study economic and social outcomes at the county level.

4 Social Connectedness and Cross-County Activity

In the previous sections, we analyzed the factors that predict the degree of cross-county social connectedness; we also documented how the geographic concentration of friendship networks correlates with important social and economic outcomes at the county level. In this section, we analyze whether the social connectedness between two regions is correlated with the degree of economic and social interaction between these regions. Specifically, we consider correlations between the number of friendship links and trade flows, patent citations, and migration patterns. As before, we focus on documenting salient patterns in the data rather than providing full-fledged causal analyses of these patterns. We view our findings as serving two purposes. First, the resulting relationships provide validation that the SCI does indeed provide a sensible measure of social connectedness at the county level. Second,

show that once we look only within states and commuting zones, those counties with more geographically concentrated social networks appear to offer greater social mobility.

[16] Likewise, Appendix Figures A5 and A6 show similar relationships as in Appendix Figure A4, but condition on the state and the commuting zone, respectively.

the patterns we document are highly consistent with many theories of an important causal role played by social interactions across a number of social and economic spheres. The results thus highlight the potential uses of the SCI data for the broader research community.

4.1 Social Connectedness and Within-U.S. Trade Flows

A well-established empirical result in the trade literature is that bilateral trade between two regions decreases with geographic distance, but the explanations for this finding are still being debated (see Anderson and van Wincoop, 2004, for a review). One proposed channel is that trade costs associated with tariffs and the transportation of goods increase with distance. However, many studies have highlighted that the distance effect is too large to be fully explained by these costs alone.[17] These papers suggest that geographic distance instead proxies for other trade frictions, such as cultural differences, lack of familiarity, or information asymmetries. In these theories, social connections may facilitate more trade if they provide a channel to alleviate the trade costs associated with information frictions. Along these lines, recent empirical work has examined the causal effect of stronger social networks on trade (see Rauch, 1999; Combes, Lafourcade and Mayer, 2005; Millimet and Osang, 2007; Cohen, Gurun and Malloy, 2012; Burchardi and Hassan, 2013; Chaney, 2014, 2016). However, much of this literature has struggled to measure the social connectedness between trading partners, and has thus had to rely on indirect proxies, such as the ethnic composition of regions or past migration patterns.

In this section, we use the SCI data to directly examine the relationship between trade flows and social connectedness at the state level. We use U.S. state-level trade flows data from the Commodity Flow Survey (CFS) to measure interstate trading volumes. We focus on data from 2012, the latest year with comprehensively available data.[18] We analyze the correlation between trade flows and social connectedness using the "gravity equation" given by regression 4.

$$\log(v_{ij}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \epsilon_{ij} \tag{4}$$

The dependent variable, $\log(v_{ij})$, captures the log of the value of trade in 2012 between origination state i and destination state j.[19] The variable $\log(d_{ij})$ denotes the log of geographic distance between states i and j,[20] 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). Other control variables, given by X_{ij}, capture differences between states i and j on measures such as GDP per capita, unemployment rates, sectoral composition,

[17] For instance, Glaeser and Kohlhase (2004) argue that the effect of distance on trade cannot be fully accounted for by shipping costs, since 80 percent of all shipments occur in industries where shipment costs are less than 4 percent of total value. See Disdier and Head (2008) for a summary of the empirically estimated distance effects in the literature.

[18] These data are collected through a survey of establishments by the U.S. 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. See the 2012 CFS Survey Methodology documentation for more information.

[19] We measure trade volume as shipment value between the originating and destination states. All patterns documented below 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). For example, column 5 of Table 3 shows that social connectedness is also a significant explanatory variable for state trade flows when we measure trade by shipment weight in tons, instead of shipment value.

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

union share, and population density.[21] 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. Standard errors are double-clustered by origin and destination states.

Column 1 of Table 3 shows the estimated elasticity of trade to geographic distance from equation 4, without controlling for social connectedness. Column 2 shows the estimated elasticity of trade to social connectedness, without controlling for geographic distance. Column 3 controls for both the log of geographic distance and the log of the SCI. Figure 7 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).

Table 3: Within-U.S. Trade and Social Connectedness

	Log(Value)				Log(Tons)
	(1)	(2)	(3)	(4)	(5)
Log(Distance)	-1.057***		-0.531***	-0.533***	-1.044***
	(0.071)		(0.084)	(0.085)	(0.101)
Log(SCI)		0.999***	0.643***	0.637***	0.768***
		(0.051)	(0.071)	(0.060)	(0.102)
State Fixed Effects	Y	Y	Y	Y	Y
Other State Differences	N	N	N	Y	Y
N	2,219	2,220	2,219	2,219	1,935
R-Squared	0.912	0.918	0.926	0.930	0.895

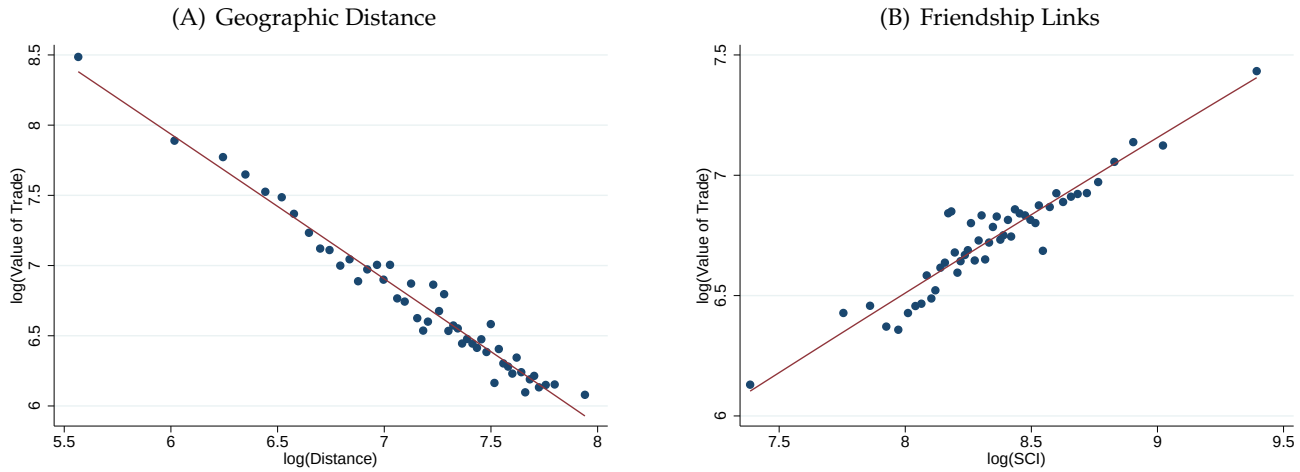
Note: Table shows results from regression 4. The unit of observation is a state-pair. The dependent variable in columns 1 through 4 is the log of the value of commodity flows between the states, and the log of the weight of commodity flows in in column 5. All specifications include fixed effects for 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. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

We observe two patterns. First, social connectedness is strongly correlated with state-state trade flows, even after controlling for geographic distance (see column 3 of Table 3, and Panel B of Figure 7). The magnitude of the elasticity of trade with social connectedness is large and statistically significant. In fact, when comparing the R^2 across columns 1 and 2, it appears as if social connectedness can explain marginally more of the variation in state-state trade flows than geographic distance.

Second, controlling for social connectedness significantly reduces the estimated distance elasticities of trade. A comparison of columns 1 and 3 shows that the distance elasticities of trade halves in magnitude after controlling for social connectedness. The coefficients are little changed when we

[21] 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.

Figure 7: State-Level Trade Flows



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.

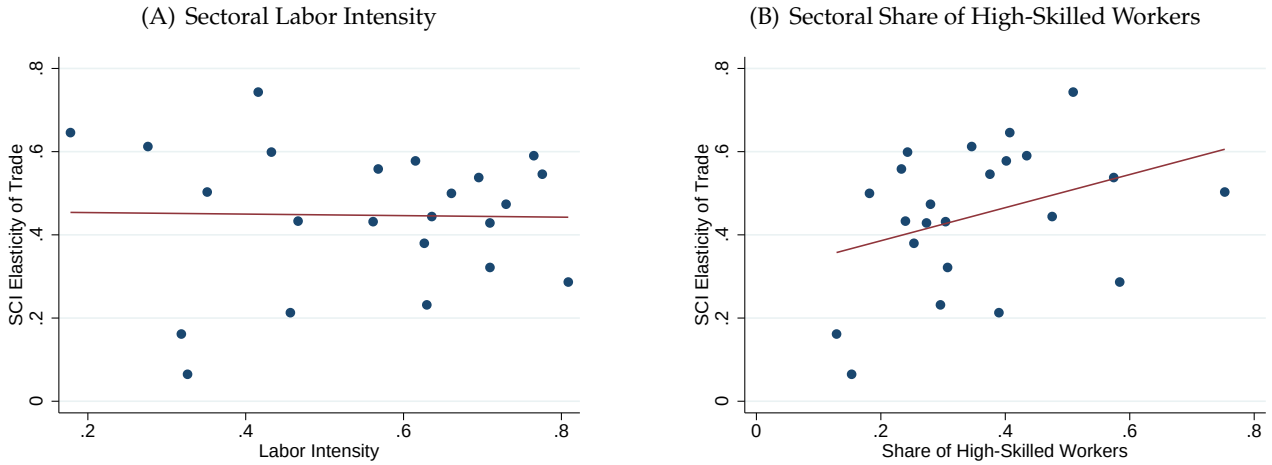
further control for other state differences in column 4. This reduction in the distance elasticities of trade, after controlling for social connectedness, supports the theories described above which suggest that geographic distance might be proxying for other factors affecting trade between states that are related to social connectedness (recall from Section 2 that geographic distance and friendship links are highly correlated). Further investigating the role of social connectedness in explaining trade flows might therefore be a useful exercise for better understanding and potentially resolving the puzzle of the high estimated geographic distance effects on trade highlighted in the literature. We hope that the availability of the SCI will help overcome some of the measurement challenges that have previously complicated such an investigation.

To understand why trade volume rises with friendship links, we further examine the variation of friendship elasticities of trade across major commodity sectors, and analyze how it varies with the labor and skill intensities of these sectors.[22] Specifically, we estimate equation 4 for each of the 23 major sectors in the CFS data. Panel A of Figure 8 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 capital. There is no discernible correlation between the elasticities and labor intensities. Panel B of Figure 8 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,

[22] 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.

with a standard error of 0.21).[23]

Figure 8: 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.

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.

4.2 Social Connectedness and Patent Citations

In many models of endogenous growth, knowledge spillovers among individuals or firms are an important driver of productivity and economic growth (Romer, 1986; Lucas, 1988; Aghion and Howitt, 1992). Social connectedness can therefore have important effects on economic activity, by facilitating the diffusion of knowledge and ideas through society.[24] However, testing the predictions from these theories is challenging, both because knowledge spillovers are hard to measure, and because of the difficulties in measuring social connectedness. To overcome these challenges, a large empirical literature has relied on patent citations as a measure of knowledge spillovers (see Jaffe, Trajtenberg and

[23] These patterns are not driven by differences in the gender compositions of each sector. We also observe a positive relationship between the friendship elasticities and the share of high-skilled male workers, and similarly for the share of high-skilled female workers.

[24] See, for example, the work of Jovanovic and Rob (1989), Kortum (1997), Benhabib and Spiegel (2005), Alvarez, Buera and Lucas. (2008), Comin and Hobijn (2010), Comin and Mestieri (2010), Comin, Dmitriev and Rossi-Hansberg (2012), Fogli and Veldkamp (2012), and Buera and Oberfield (2016). Social networks can also affect the exposure of the region to new ideas and therefore influence how quickly the region adopts a new idea. See for instance Glaeser (1999), Black and Henderson (1999), and Moretti (2012) for studies on the role of geography in shaping innovation outcomes.

Henderson, 1993; Thompson and Fox-Kean, 2005). By studying the geographic distances between the locations where the issued patents and patent citations occur, these papers conclude that knowledge spillovers are highly localized. This, in turn, is often interpreted as evidence for the role of social interactions, which are more likely to happen at shorter distances. Other attempts to measure social connectedness have tried to proxy for an inventor's peer group based on characteristics such as common ethnicity (Agrawal, Kapur and McHale, 2008). In this section, we show that through measuring social connectedness via the SCI, we can advance upon the existing literature, and provide more direct evidence for the role of social connectedness in facilitating knowledge spillovers.

Our data contain information on all patents granted by the USPTO in the years 2002-2014, as well as information on the location associated with the patent. This location is based on the location of the company or institution from which the patent originated. If the company or institution is not available, then the patent is assigned to the location of the first inventor with an available location.[25] The patents cover 107 different technological classes, defined based on the International Patent Classification. For each granted patent, we observe all other patents that it cites.

The empirical challenge is to separate knowledge spillovers from correlations that might be induced by patterns in the geographic location of technologically related activities across regions that are connected through social networks. For example, imagine that Austin, TX, and the Bay Area have a high degree of social connectedness, perhaps because tech workers in both regions know each other from graduate school. If we see a higher incidence of patent citations across these two regions, this could either be because of knowledge spillovers along these social networks, or because tech patents are more likely to cite other tech patents. To address such concerns, we follow the approach in the existing literature to identify the causal effect of social connectedness on patent citations (see, for example, Jaffe, Trajtenberg and Henderson, 1993; Thompson and Fox-Kean, 2005; Agrawal, Kapur and McHale, 2008). This approach matches each citing patent with a non-citing "control" patent issued at the same time and in the same technological class. Knowledge spillovers are then measured as the extent to which the citation probability increases with the social connectedness of the geographies associated with the patents, over and beyond what is expected based on the technological class of the issued patent or geographic distance.

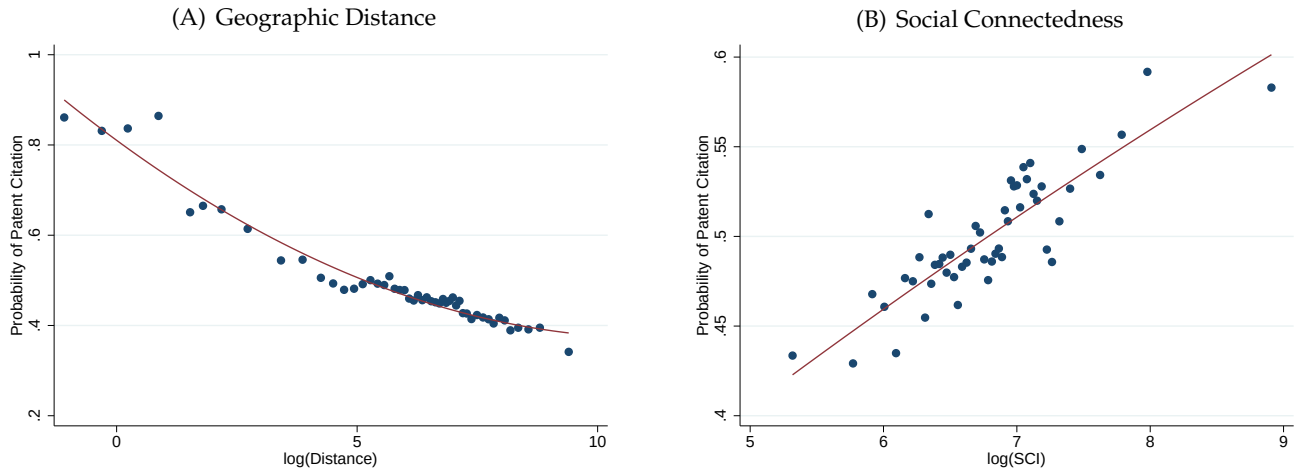
We start with all patents granted in the U.S. in 2014. For each of these 2014 patents, we create a separate observation for each of the patents cited by the 2014 patent, so that the unit of observation is a patent-citation pair. For example, if a particular 2014 patent cites 10 other patents, this will 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. Our final sample includes over 1.5 million matched patent-citation pairs.

Panel A of Figure 9 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.

[25] We thank Enrico Berkes and Ruben Gaetani for sharing their geo-referenced data set. See Berkes and Gaetani (2016) for a more detailed discussion of the data.

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.

Figure 9: 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.

In Panel B of Figure 9, 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.

We further examine the relationship between friendship links and patent citations using equation 5. 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 the 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, denoted by $\psi_{s(i)}$ and $\psi_{s(j)}$ respectively. We double cluster the standard errors by the technology classes of patents i and j.

$$C_{ij} = \beta_1 \log(d_{c(i)c(j)}) + \beta_3 \log(f_{c(i)c(j)}) + \beta_3 X_{c(i)c(j)} + \psi_{c(i)} + \psi_{c(j)} + \psi_{s(i)} + \psi_{s(j)} + \epsilon_{ij} \quad (5)$$

Column 1 of Table 4 shows the effect of distance on the probability of citation from equation 5, but without controlling for the degree of social connectedness and other across-county differences. We

find that the probability of citation falls by a statistically-significant 4.8 percentage points if the distance between the counties of the issued and cited patents doubles.[26]

Table 4: Patent Citations and Social Connectedness

	(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

Note: Table shows results from regression 5. 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. See text for more details. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Column 2 of Table 4 shows the effect of social connectedness on the probability of citation, without controlling for distance and other across-county differences. We estimate that the probability of citation rises by 6.3 percentage points if the social connectedness between the counties of the issued and cited patents doubles. The effect is statistically significant. All else equal, social connectedness explains more of the variation in the probability of a patent citation than geographic distance.

In Column 3 of Table 4, we jointly estimate the effects of distance and social connectedness on the probability of citation. We find that the effect of doubling social connectedness on the probability of citation remains significant and large, at 4.9 percent, even after controlling for geographic distance. In comparison, the effect of doubling geographic distance on the probability of citations falls from -4.8 percent to -1.1 percent. This suggests that the distance variable may be primarily capturing information flows associated with social connectedness.

In columns 4 to 6 of Table 4, we vary the controls that are included in the regression. Our estimates change little when we further control for other across-county differences in column 4. In column 5, we include fixed effects for the cited patent, and in column 6, we include fixed effects for the issued and cited patents. In all specifications, we observe a statistically significant and positive effect of social connectedness on the probability of citation.

Overall, these findings show that the SCI data allow us to detect a significant correlation between social connectedness and innovative activity. We hope that these initial results will encourage other

[26] Previous studies, such as Agrawal, Kapur and McHale (2008) estimate that a 1000-mile increase in distance reduces the probability of citation by approximately 2 percentage points. The relatively low R-squared for the regressions in Table 4 are comparable to past studies.

researchers to use the SCI data to better understand the effects of social interactions on the relationship between knowledge spillovers, innovation, and economic growth.

4.3 Social Connectedness, Migration, and Labor Market Flows

We next analyze the extent to which the social connectedness between two regions affects the flow of people across these regions. Understanding the factors driving migration patterns is important, for example because within-U.S. migration is integral to equilibrating the U.S. labor market following regional shocks (Blanchard and Katz, 1992). Despite a large body of research analyzing various aspects of U.S. migration, many important aspects of migration patterns, such as the decline in geographic mobility in the U.S. since the 1980s, are not fully understood (see, for example, Molloy, Smith and Wozniak, 2011; Kaplan and Schulhofer-Wohl, 2012).

An existing literature has documented that social networks can play an important role in facilitating migration, by providing information as well as social and economic support (see Munshi, 2014, for a review). While a lot of the research has focused on international migration, similar forces might be at work in explaining within-U.S. migration. To highlight the potential for the SCI to advance our understanding of the role of social networks in facilitating within-U.S. migration, we next show that differences in social connectedness have significant explanatory power for migration and labor market flows between regions, beyond what is predicted by geographic distance. The results suggest an important role for social networks in shaping population and labor market flows within the U.S., consistent with across-country migration studies such as Moretti (1999). This evidence might help to calibrate models of migration flows such as that in Carrington, Detragiache and Vishwanath (1996).

County-County Population Flows. We first 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. As a first piece of evidence that the number of movers is linked to the degree of social connectedness, the 95^{th} percentile of the SCI among those county-pairs with fewer than 20 movers is below the 5^{th} percentile of the SCI among county-pairs that have movers. The following analysis will focus on those county-pairs for which we observe the number of movers.

We analyze the relationship between social connectedness and population flows using the specification in equation 6. 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.

$$\log(m_{ij}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \epsilon_{ij} \tag{6}$$

In column 1 of Table 5 we do not include the control for the social connectedness of the two counties. The estimated elasticity of migration to geographic distance is close to -1. Panel A of Figure 10 shows a binned scatter plot that documents this relationship non-parametrically, controlling for county fixed effects. The relationship is not perfectly linear, with a somewhat more negative elasticity at shorter distances. In column 2 we control for geographic distance flexibly in the same way as in the binned scatter plot, by including 50 indicator variables for equally-sized parts of the distance distribution; the R^2 increases somewhat, due to the flexible controls' ability to better approximate the non-linear relationship.

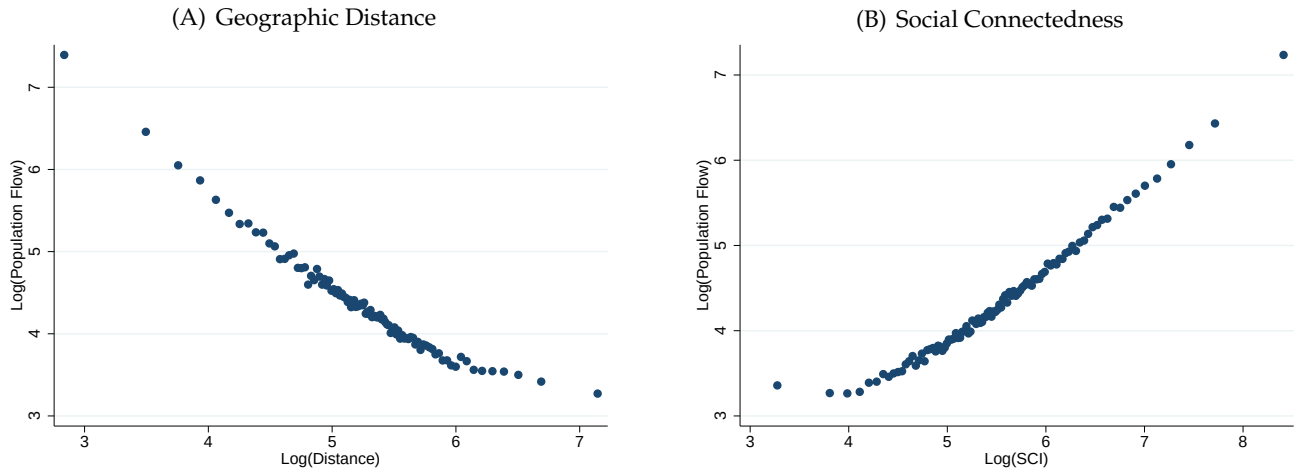
Table 5: Migration and Social Connectedness

	(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: Table shows results from regression 6. 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. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

In column 3 of Table 5 we do not control for distance, but instead include the control for the log of the SCI. The elasticity of migration to social connectedness is slightly larger than 1. Importantly, the R^2 is substantially higher than when controlling flexibly for distance – this suggests that the SCI can explain a larger part of the variation of the migration flows across county-pairs than geographic distance can. In columns 4 and 5 we also control for the geographic distance between county-pairs. This control variable has no additional predictive power, and the R^2 hardly increases. This finding suggests that much of the effect of distance on migration might be coming from the relationship between distance and social connectedness, and that distance by itself has no additional explanatory power for migration. Panel B of Figure 10 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 linear, suggesting a constant elasticity between friendship links and migration. Finally, in column 6 of Table 5, we also control for differences across the county-pairs in other characteristics such as income, education levels, race, and voting patterns. The inclusion of these additional controls does not have a significant effect on R^2, suggesting that any predictive power of these variables is already captured by our measure of social connectedness.

Figure 10: 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.

Overall, our results suggest that individuals are much more likely to move to counties where they already have friends. This induces a force that means that cities that are already large continue to attract more and more new people; such a force might help explain the very right-tailed city size distribution (Gabaix, 1999).

State-State Job Flows. A second data set to analyze within-U.S. migration comes from the U.S. 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 U.S. 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 7 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 $\log(f_{ij})$ denotes the log of the relative number of friendship links (i.e., 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. We double cluster the standard errors at the levels of the origin and destination states.

$$\log(y_{ijt}) = \beta_1 \log(d_{ij}) + \beta_2 \log(f_{ij}) + \beta_3 X_{ij} + \psi_i + \psi_j + \psi_t + \epsilon_{ijt} \tag{7}$$

Table 6 reports the estimates from equation 7, focusing on the within-quarter job flows.[27] Column 1 does not control for the SCI. The estimated elasticity of job flows with respect to geographic distance is very close to -1, resembling the migration elasticity estimates from the county-county migration data. Panel A in Figure 11 shows a binned scatter plot that documents the nearly linear relationship non-parametrically, controlling for state fixed effects.

Table 6: Labor Flows and Social Connectedness

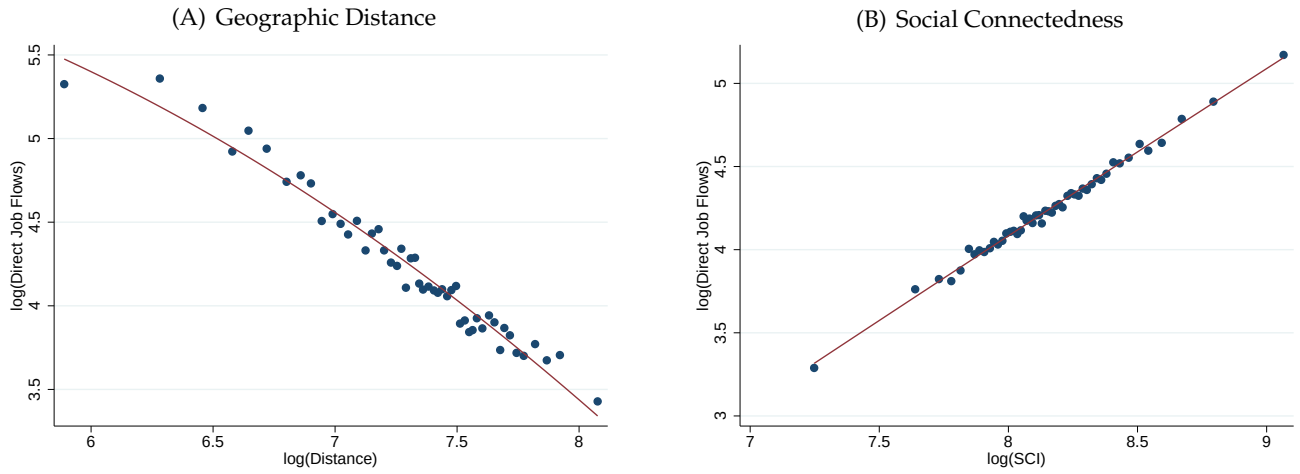
	(1)	(2)	(3)	(4)	(5)
Log(Distance)	-1.001***		-0.221***	-0.209***	-0.208***
	(0.062)		(0.022)	(0.022)	(0.022)
Log(SCI)		1.152***	1.007***	1.000***	1.001***
		(0.025)	(0.027)	(0.028)	(0.030)
Other State Differences	N	N	N	Y	Y
State and Time Fixed Effects	Y	Y	Y	Y	N
State x Time Fixed Effects	N	N	N	N	Y
N	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 7. 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: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

In column 2 of Table 6, 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 Figure 11 shows a binned scatter plot that documents the strong and nearly linear relationship between the logs of 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, 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 Table 6, 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

[27] 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.

Figure 11: State-Level Labor Flows



Note: Both panels show the estimated friendship elasticity of labor flows from equation 7. 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.

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.

4.4 Social Connectedness and the Spread of Sentiments

Social networks can be an important source for the transmission of sentiments such as feelings of general optimism or pessimism. For example, Shiller (2007) writes that "many people seem to be accepting that the recent home price experience is at least in part the result of a social epidemic of optimism for real estate." Social dynamics of optimism about house price growth also play an important role in the narrative of housing booms and busts in Burnside, Eichenbaum and Rebelo (2015), and have been hypothesized to play a role in understanding the geographic spread of house prices documented by DeFusco et al. (2015). Sentiments that spread through social networks can also help explain other economic phenomena, such as cyclical swings in economic activity (Blanchard, L'Huillier and Lorenzoni, 2013; Angeletos and La'O, 2013).

Despite the importance of social networks in propagating sentiments through the economy, empirical analyses have been complicated by the absence of suitable data to analyze these mechanisms. The SCI data provides new opportunities for such important empirical work. For example, Bailey et al. (2016*a*) document that recent house price experiences within individuals' social networks affect their perceptions of the attractiveness of property investments, and through this channel have large effects on their housing market activity. In related work, Bailey et al. (2016*b*) use data similar to the SCI introduced in this paper to show that house price increases in some counties lead to higher house price growth in other geographically distant counties that are connected through friendship

links. This house price effect arises over and above what would be predicted by common shocks to connected counties.

4.5 Additional Potential Applications of the SCI Data

In addition to the settings that we explore in this section, there are numerous other research and policy questions that can be pursued with the county-level SCI data:

1. Many contagious illnesses and diseases, such as the flu or tuberculosis, spread through human contact. Combined with localized data on the prevalence of the flu, data on social connectedness might therefore allow researchers and public health officials to better predict where to expect future outbreaks of the flu (see Cauchemez et al., 2011; Christakis and Fowler, 2010; Glass and Glass, 2008).

2. While we discuss research that suggests that house price sentiments spread through social networks, the SCI could also be used to track whether other measures of sentiment, for example those tracked by the Michigan Survey of Consumers, or through geo-coded Twitter feeds, have similar geographic patterns.

3. Sociolinguistic research has argued that social networks are an important force determining linguistic development (e.g., Milroy, 1987). The SCI data would allow researchers to empirically study the extent to which linguistic development in the U.S. can be explained by patterns of social connectedness.

4. Given today's highly diversified media environment (see New York Times, 2016), targeting advertising to reach specific demographics is now more effective than targeting entire media markets. With the SCI data, it is possible to identify geographically distant regions with strong connectivity, which, given the strong evidence for homophily in determining friendship linkages, may indicate similarities along numerous demographic dimensions. This can be used to study the effectiveness of advertising using data on product adoption in areas targeted by different product lines, or for more effectively directing advertising strategies.

5. Significant social connectedness between two regions might be a strong indicator that providing transportation infrastructure between these regions, such as direct airline routes, might be profitable. Using the SCI as a measure of the potential demand for various routes could solve some of the identification issues in the literature analyzing airline scheduling in operations research and industrial organization (e.g., Molnar, 2013).

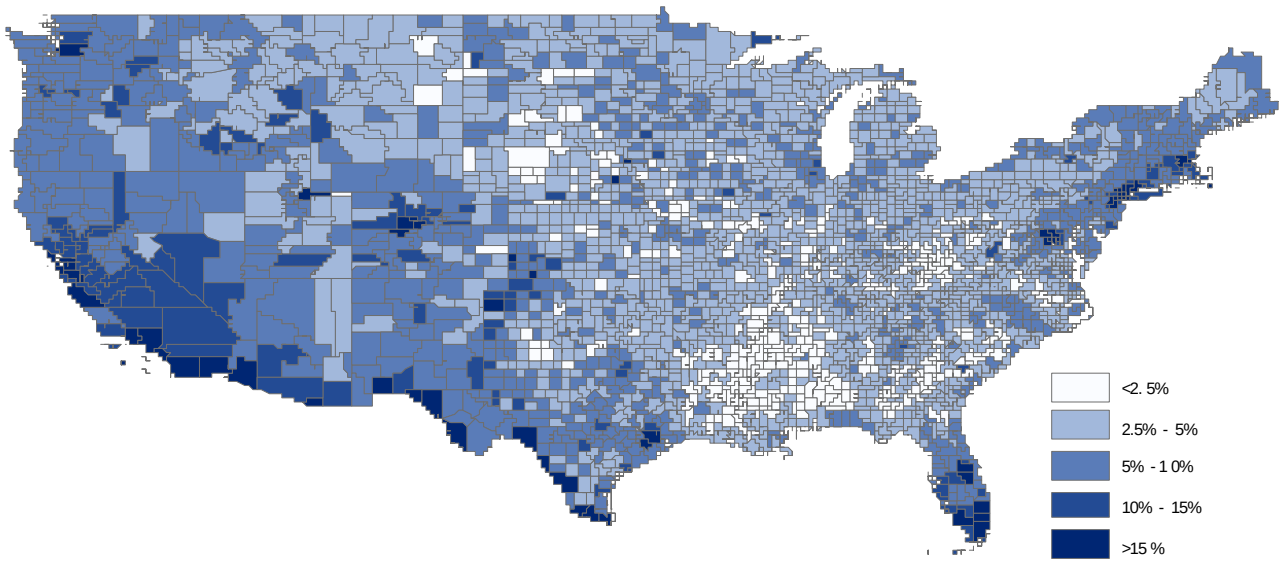
6. Relatedly, while strong social connectedness between regions may provide reasons for expanding transportation links between those regions, increased transportation links can also have a causal effect on social connectedness. The SCI data is ideal for analyzing this relationship. One approach would be to compare the social connectness of two counties that happen to lie on the straight line between two major cities, and which are therefore connected by a highway, to the connectness of two similar counties that do not lie on the straight line between major cities.

5 International Dimension of Social Connectedness of U.S. Counties

In this section, we explore the social connectedness between U.S. counties and foreign countries. We focus on how today's social connectedness is correlated with (i) past migration patterns, and (ii) present-day economic activity.

We begin by exploring summary statistics of the international connectedness of U.S. counties. There is large heterogeneity across U.S. counties in the share of social connections to individuals living outside of the United States. The median county has 4% of all friendship links to individuals living in foreign countries, but the 10-90 percentile range is 2.3% to 8.6%, and the 1-99 percentile range is 1.6% to 18.7%. Figure 12 shows a heatmap of "non-U.S. friend" shares across counties in the United States; unsurprisingly, regions that are close to U.S. land borders have more friendship links with foreign countries. Similarly, Florida and the Northeast have more friendship links to individuals living outside of the United States.

Figure 12: Social Connectedness Abroad



Note: Figure shows a heatmap of the share of friendship links that are to Facebook users outside of the United States.

In Figure 13, we explore the relative share of friendship links of counties in the continental U.S. to six foreign countries: Canada, Mexico, Germany, Italy, Finland, and Somalia. For both Mexico and Canada, there are significantly stronger links to those counties close to a land border with the country. Relative connections with Germany are particularly strong for those counties in the Midwest (particularly in Michigan, Wisconsin, and Minnesota) and on the West Coast that saw major immigration from Germany in the late 19th Century.[28] Similarly, present-day social connectedness with Italy is particularly strong in those Northeastern counties that experienced substantial Italian immigration in the late 19th and early 20th Centuries. The SCI also allows us to identify present-day links to relatively small countries with more limited migration to the United States. Panel E shows linkages to Norway,

[28]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.

which are mostly concentrated in the upper Midwest, a region that is home to most Americans of Norwegian descent (see Wikipedia, 2005, for a map showing the geographic distribution of Norwegian Americans). Similarly, 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.[29]

5.1 Immigration and Social Connectedness

We next analyze more formally 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. The second measure is broader and can, for instance, include U.S.-born individuals with immigrant parents or grandparents. 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 U.S. claim ancestry.[30] 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. In some of our analysis, we will distinguish countries by when immigration to the U.S. 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 14 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.

Next, we estimate the following regression equation to formally analyze the effect of past migration on current social connectedness for all countries in our data:

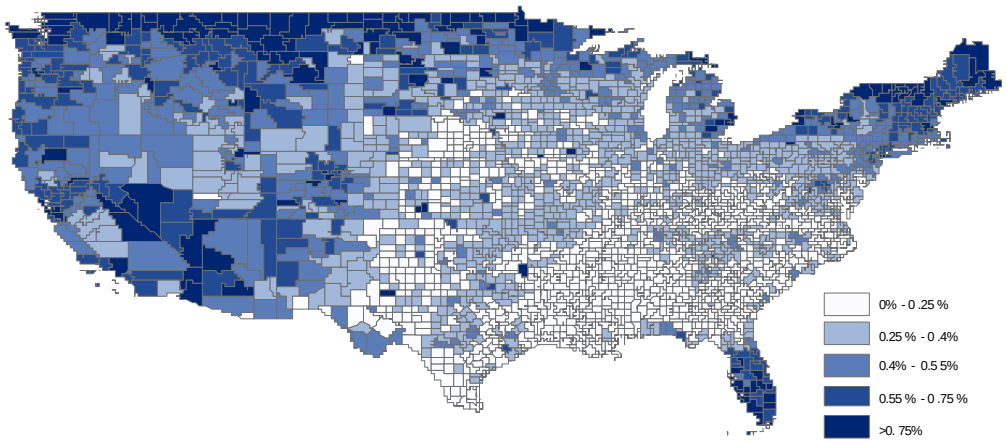
$$log(f_{ic}) = \beta_1 log(anc_{ic}) + \beta_2 log(d_{ic}) + \psi_c + \psi_i + \epsilon_{ic}, \tag{8}$$

[29] Importantly, while we can compare the relative connectedness of a particular foreign country with different U.S. counties, it is harder to interpret the relative connectedness across two different countries. This is because the Facebook penetration differs across countries. This means that a particular county might have more Facebook friendship links with country A than with country B because the true social connectedness with country A is higher, or because there are (relatively) more Facebook users in country A.

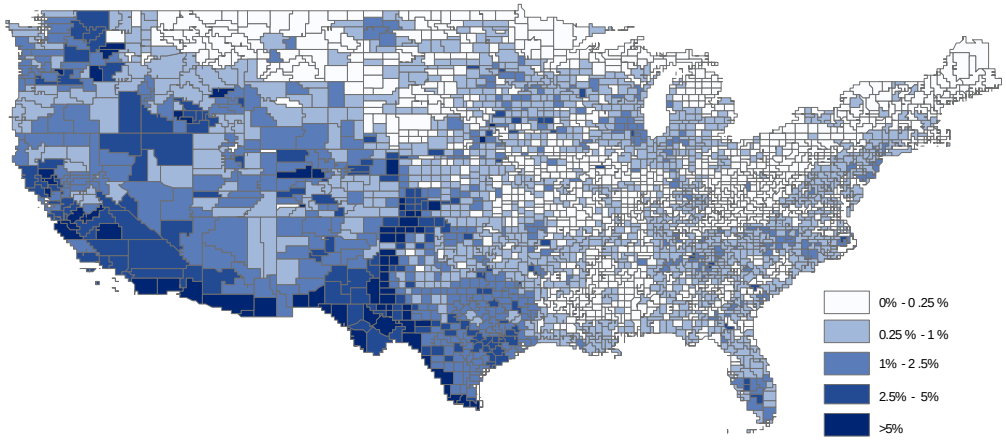
[30] Our results are robust to varying this cut-off or including all countries. We also re-code ancestry and foreign birthplaces to match today's political boundaries, such as coding "Persian" ancestry to "Iran."

Figure 13: Ancestry and Social Connectedness

(A) Canada



(B) Mexico



(C) Germany

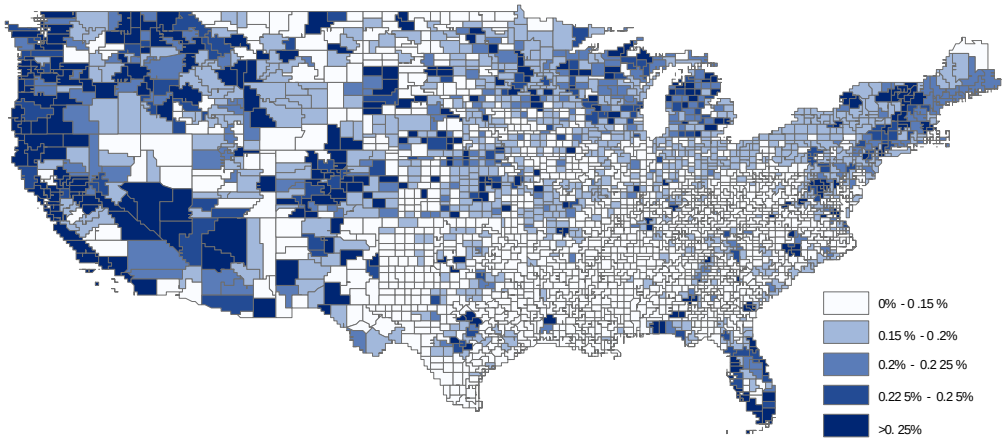
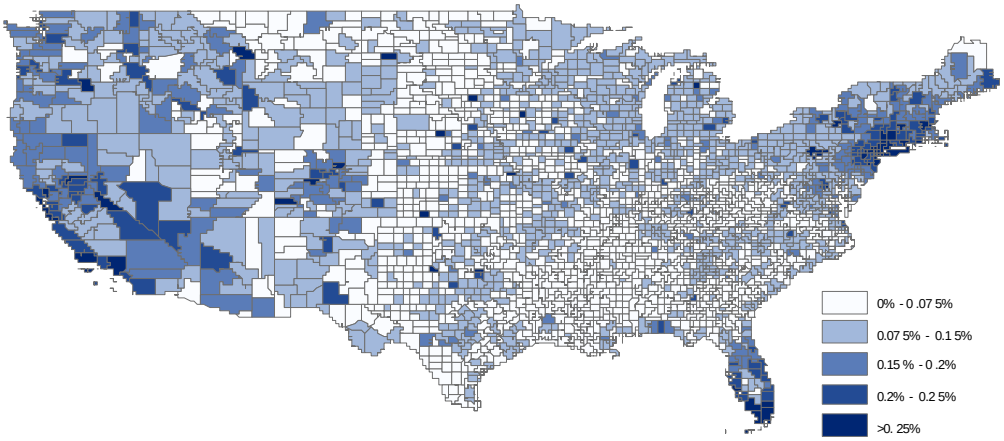
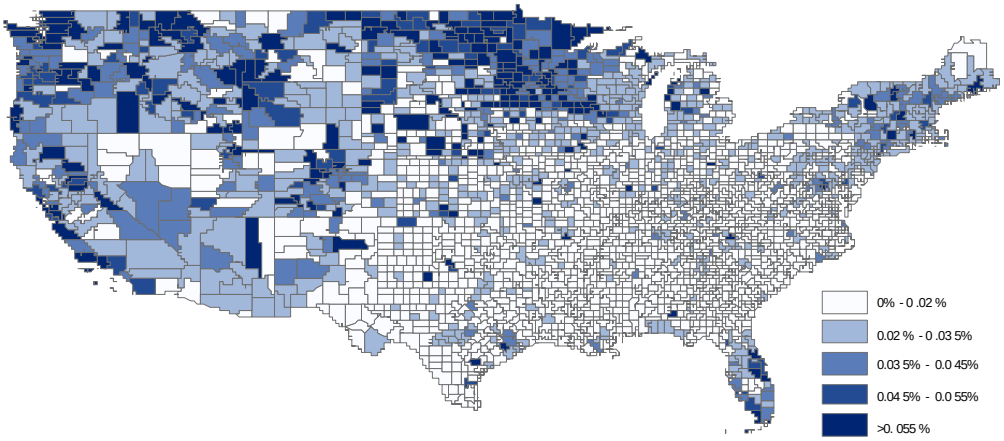


Figure 13: Ancestry and Social Connectedness

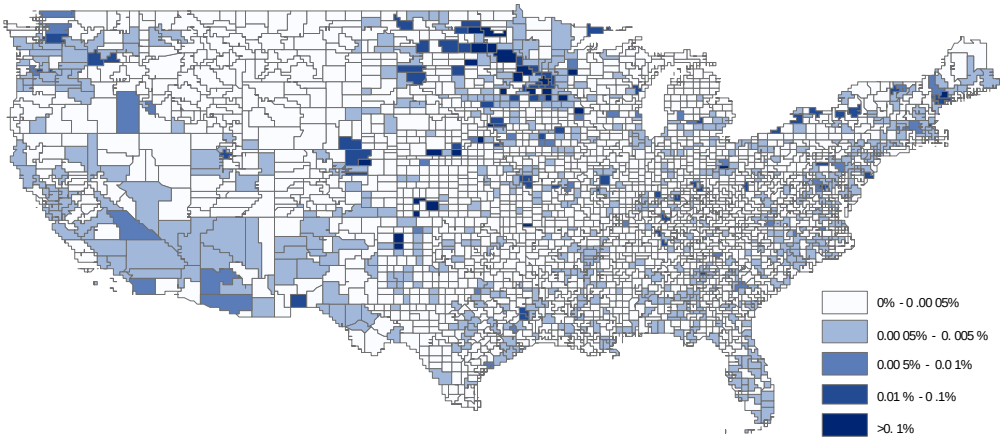
(D) Italy



(E) Norway

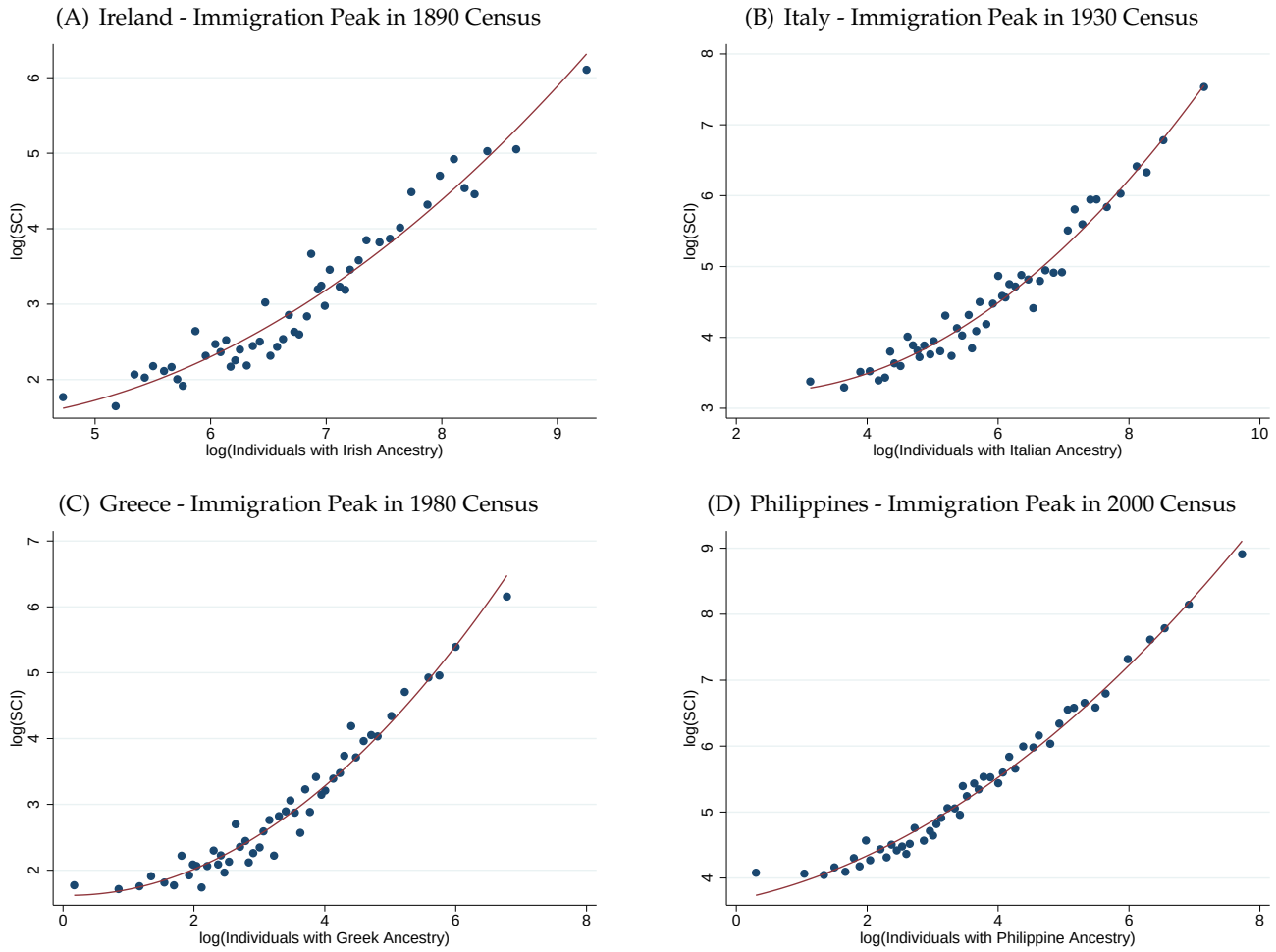


(F) Somalia



Note: Figure shows a heatmap of the share of friendship links that are to Facebook users located in Canada (Panel A), Mexico (Panel B), Germany (Panel C), Italy (Panel D), Norway (Panel E), and Somalia (Panel F).

Figure 14: Ancestry and Social Connectedness



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.

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, we also estimate the above equation with $log(f_{ic} + 1)$ as the dependent variable; in those specifications, the key explanatory variable is $log(anc_{ic} + 1)$.

Table 7 displays the results. The first column shows the effect of geographic distance on today's social connections: a one percent increase in the geographic distance is associated with a 1.2% decline in social connectedness. Interestingly, this elasticity is nearly identical to the elasticity of friendship links to geographic distance estimated for the U.S. for distances greater than 200 miles. The remaining columns include measures of past migration as additional controls. Past migration has a substantial effect on today's social connectedness. A 1% percent increase in the number of residents with ancestry from a given foreign country increases social connections to that country by about a third of a percent.

Table 7: 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

Note: Table shows results from regression 8. The unit of observation is a U.S. county-foreign country pair. Each specification also includes fixed effects for the U.S. state and the foreign country. Standard errors are clustered at both the county and foreign country level. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

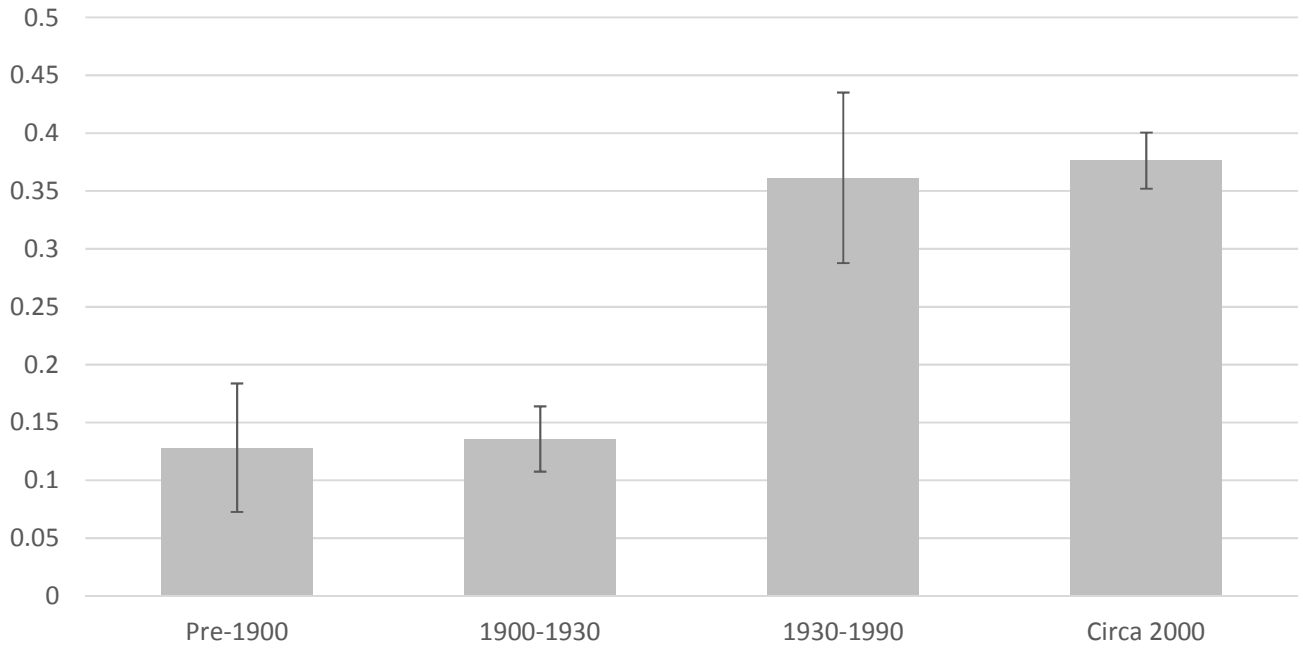
Adding past migration also reduces the effect of distance by between a third and a half. The estimates are of similar magnitude when we focus only on respondents born in a given foreign country. They are also similar when using the $log(f_{ic}+1)$ instead of the $log(f_{ic})$ specifications, indicating that including the substantial number of county-country pairs with zero ancestry or friendship links does not alter our results. Finally, in unreported results, we find very similar estimates when using ancestry data from the 2000 Census instead of the 2014 5-year ACS.

Next, we analyze whether the effect of past migration on today's social connections is stronger for countries from which immigration to the U.S. occurred more recently, such as Mexico or the Philippines, compared to countries from which immigration peaked earlier, such as Germany or Ireland. Figure 15 shows the effect of ancestry on current social connections by year of peak immigration. Ancestry has a substantially larger effect on today's social connections if immigration peaked after the Second World War compared to countries with earlier immigration peaks. However, even past immigration from countries with few migrants for several generations, such as Germany and Ireland, predicts higher social connectedness today. These estimates allow researchers to parameterize how social connectedness of migrants with their home countries decays over time.

5.2 Social connections and international trade

In this section, we analyze the extent to which higher social connectedness with foreign countries is associated with more trade with these countries. This mirrors the analyses in Section 4.1, which estimated the role of social connectedness in facilitating within-U.S. trade. This analysis contributes to a

Figure 15: Ancestry and Social Connectedness by Migration Peak



Note: Figure shows estimates of β_1 from regression 8, separately for countries grouped by the census year of peak immigration.

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 (see our discussion in Section 4.1); 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 U.S. 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 international 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 U.S. dollars. Specifically, we estimate the relationship between today's social connectedness and imports and exports with the following regression:

$$log(t_{is}) = \beta_1 log(f_{is}) + \beta_2 log(d_{is}) + \psi_s + \psi_i + \epsilon_{is}, \tag{9}$$

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 $log(f_{is}+1)$.

Table 8: Social Connectivity 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: Table shows results from regression 9 for 2015 levels of total imports (columns 1 to 3) and 2015 levels of exports (column 4 - 6), both in current U.S. Dollars. The unit of observation is a U.S. state-foreign country pair. Each column also includes fixed effects for the U.S. state and for the foreign country. Standard errors are clustered at both the state and foreign country level.

Table 8 shows that including the SCI as an explanatory variable reduces the effect of distance on international trade substantially, by about one third for imports and a quarter for exports. Social connectedness itself strongly affects 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 9 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.

6 Conclusion

We use data from the world's largest online social networking site, Facebook, to construct a Social Connectedness Index (SCI). The SCI provides a new and comprehensive measure of social connectedness between U.S. county pairs, as well as between U.S. counties and foreign countries. We argue that these data allow researchers to overcome many of the measurement challenges that have held back empirical research on the role of social interactions in finance, economics, and the broader social sciences. To illustrate this point, we show how the SCI data can be used both to better understand the geographic dimensions of real-world social networks, as well as to highlight that social interactions correlate with social and economic activity across regions. For example, we document a strong relationship between social connectedness and trade, both across U.S. states and between U.S. states and foreign countries. While not all of these correlations necessarily identify a causal relationship, they provide starting points for research that can build on the SCI in order to address a wide variety of questions of academic and policy interest.

Bibliography

Aghion, P, and P Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica*, 60(2).

Agrawal, Ajay, Devesh Kapur, and John McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics*, 64: 258–269.

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas. 2008. "Models of Idea Flows." National Bureau of Economic Research Working Paper 14135.

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

Angeletos, George-Marios, and Jennifer La'O. 2013. "Sentiments." *Econometrica*, 81(2): 739–779.

Bailey, Michael, Eduardo Davila, Theresa Kuchler, and Johannes Stroebel. 2017. "House Price Beliefs and Leverage."

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2016*a*. "Social Networks and Housing Markets." *Working Paper*.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, Johannes Stroebel, and Joseph Vavra. 2016*b*. "Prices and Trading Volumes in the Housing Markets: The Role of Disagreements and Social Networks." *Working Paper*.

Benhabib, Jess, and Mark Spiegel. 2005. "Human Capital and Technology Diffusion." In *Handbook of Economic Growth*. Vol. 1, Part A. 1 ed., , ed. Philippe Aghion and Steven Durlauf, Chapter 13, 935–966. Elsevier.

Berkes, Enrico, and Ruben Gaetani. 2016. "The Geography of Unconventional Innovation." *Unpublished Manuscript*.

Black, Duncan, and J. Vernon Henderson. 1999. "A Theory of Urban Growth." *Journal of Political Economy*, 107(2): 252–284.

Blanchard, Olivier Jean, and Lawrence F Katz. 1992. "Regional evolutions." *Brookings papers on economic activity*, 1992(1): 1–75.

Blanchard, Olivier J, Jean-Paul L'Huillier, and Guido Lorenzoni. 2013. "News, noise, and fluctuations: An empirical exploration." *The American Economic Review*, 103(7): 3045–3070.

Blondel, Vincent, Gautier Krings, Isabelle Thomas, et al. 2010. "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone." *Brussels Studies*.

Bramoulle, Yann, Andrea Galeotti, and Brian Rogers. 2016. *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Buera, Francisco J., and Ezra Oberfield. 2016. "The Global Diffusion of Ideas." National Bureau of Economic Research Working Paper 21844.

Burchardi, Konrad B., and Tarek A. Hassan. 2013. "The Economic Impact of Social Ties: Evidence from German Reunification." *The Quarterly Journal of Economics*, 128(3): 1219–1271.

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

Burnside, A Craig, Martin Eichenbaum, and Sergio Rebelo. 2015. "Understanding booms and busts in housing markets." *Journal of Political Economy*.

Calabrese, Francesco, Dominik Dahlem, Alexandre Gerber, DeDe Paul, Xiaoji Chen, James Rowland, Christopher Rath, and Carlo Ratti. 2011. "The connected states of america: Quantifying social radii of influence." 223–230, IEEE.

Carrington, William J, Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with endogenous moving costs." *The American Economic Review*, 909–930.

Cauchemez, Simon, Achuyt Bhattarai, Tiffany L Marchbanks, Ryan P Fagan, Stephen Ostroff, Neil M Ferguson, David Swerdlow, Samir V Sodha, Mària E Moll, Frederick J Angulo, et al. 2011. "Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza." *Proceedings of the National Academy of Sciences*, 108(7): 2825–2830.

Chaney, Thomas. 2014. "The Network Structure of International Trade." *American Economic Review*, 104(11): 3600–3634.

Chaney, Thomas. 2016. *Oxford Handbook of the Economics of Networks.* Oxford University Press.

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.

Christakis, Nicholas A, and James H Fowler. 2010. "Social network sensors for early detection of contagious outbreaks." *PloS one*, 5(9): e12948.

Cohen, Lauren, Umit G. Gurun, and Christopher J. Malloy. 2012. "Resident Networks and Firm Trade." National Bureau of Economic Research Working Paper 18312.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer. 2005. "The trade-creating effects of business and social networks: evidence from France." *Journal of International Economics*, 66(1): 1–29.

Comin, Diego A., and Martï£¡ Mestieri. 2010. "An Intensive Exploration of Technology Diffusion." National Bureau of Economic Research Working Paper 16379.

Comin, Diego A., Mikhail Dmitriev, and Esteban Rossi-Hansberg. 2012. "The Spatial Diffusion of Technology." National Bureau of Economic Research Working Paper 18534.

Comin, Diego, and Bart Hobijn. 2010. "An Exploration of Technology Diffusion." *American Economic Review*, 100(5): 2031–2059.

DeFusco, Anthony A, Wenjie Ding, Fernando V Ferreira, and Joseph Gyourko. 2015. "The Role of Contagion in the American Housing Boom."

Denver Post. 2010. "Go west for fruitful foray to 'Other Colorado'." http://extras.denverpost.com/rec/travel36.htm, accessed: 2016-12-21.

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.

Disdier, Anne-Celia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics*, 90(1): 37–48.

Duggan, Maeve, Nicole B Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. 2015. "Social media update 2014. Pew Research Center."

Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network diversity and economic development." *Science*, 328(5981): 1029–1031.

Facebook. 2016. "Facebook Form 10-Q, Quarter 3, 2016."

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

Fogli, Alessandra, and Laura Veldkamp. 2012. "Germs, Social Networks and Growth." National Bureau of Economic Research Working Paper 18470.

Gabaix, Xavier. 1999. "Zipf's law for cities: an explanation." *The Quarterly journal of economics*, 114(3): 739–767.

Gilbert, Eric, and Karrie Karahalios. 2009. "Predicting tie strength with social media." 211–220, ACM.

Glaeser, Edward. 1999. "Learning in Cities." *Journal of Urban Economics*, 46(2): 254–277.

Glaeser, E. L., and J. E. Kohlhase. 2004. "Cities, Regions and the Decline of Transport Costs." *Papers in Regional Science*, 83(1): 197–228.

Glass, Laura M, and Robert J Glass. 2008. "Social contact networks for the spread of pandemic influenza in children and teenagers." *BMC public health*, 8(1): 1.

Golub, Benjamin, and Matthew Jackson. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *The Quarterly Journal of Economics*, 127(3): 1287–1338.

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.

Granovetter, Mark. 2005. "The impact of social structure on economic outcomes." *The Journal of Economic Perspectives*, 19(1): 33–50.

Granovetter, Mark S. 1973. "The strength of weak ties." *American journal of sociology*, 1360–1380.

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

Hampton, Keith, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. "Social networking sites and our lives. Pew internet and american life project."

Holahan, CJ, BL Wilcox, MA Burnam, and RE Culler. 1978. "Social satisfaction and friendship formation as a function of floor level in high-rise student housing."

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.

Jackson, Matthew O. 2014. "Networks in the understanding of economic behaviors." *The Journal of Economic Perspectives*, 28(4): 3–22.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics*, 108(3): 577–598.

Jones, Jason J, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow, and James H Fowler. 2013. "Inferring tie strength from online directed behavior." *PloS ONE*, 8(1): e52168.

Jovanovic, Boyan, and Rafael Rob. 1989. "The Growth and Diffusion of Knowledge." *The Review of Economic Studies*, 56(4): 569–582.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2012. "Understanding the long-run decline in interstate migration." National Bureau of Economic Research.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica*, 65(6): 1389–1419.

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.

Lazarsfeld, P., and R. K. Merton. 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis." In *Freedom and Control in Modern Society*. , ed. M. Berger, T. Abel and C. H. Page. New York:Van Nostrand.

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.

Los Angeles Times. 2016. "Oil's latest slump takes a heavy toll on Bakersfield." http://www.latimes.com/business/la-fi-bakersfield-oil-20160207-story.html, accessed: 2016-12-21.

Lucas, Robert. 1988. "On the mechanics of economic development." *Journal of Monetary Economics*, 22(1): 3–42.

Marmaros, David, and Bruce Sacerdote. 2006. "How do friendships form?" *The Quarterly Journal of Economics*, 79–119.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 415–444.

Millimet, Daniel L, and Thomas Osang. 2007. "Do state borders matter for US intranational trade? The role of history and internal migration." *Canadian Journal of Economics*, 40(1): 93–126.

Milroy, Lesley. 1987. *Language and social networks.* Wiley-Blackwell.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *The Journal of Economic Perspectives*, 25(3): 173–196.

Molnar, Alejandro. 2013. "Congesting the commons: A test for strategic congestion externalities in the airline industry." Working paper.

Moretti, Enrico. 1999. "Social networks and migrations: Italy 1876-1913." *International Migration Review*, 640–657.

Moretti, Enrico. 2012. *The New Geography of Jobs.* Houghton Mifflin Harcourt.

Munshi, Kaivan. 2014. "Community Networks and Migration."

News OK. 2015. "Read this, y'all: Is the Okie dialect disappearing?" http://newsok.com/article/5420122, accessed: 2016-12-21.

New York Times. 2016. "Duck Dynasty vs. Modern Family: 50 Maps of the U.S. Cultural Divide." http://www.nytimes.com/interactive/2016/12/26/upshot/duck-dynasty-vs-modern-family-television-maps.html?_r=0, accessed: 2017-01-06.

Page, Scott E. 2008. *The difference: How the power of diversity creates better groups, firms, schools, and societies.* Princeton University Press.

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

Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H Strogatz. 2010. "Redrawing the map of Great Britain from a network of human interactions." *PloS one*, 5(12): e14248.

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

Rauch, JE. 1999. "Networks versus markets in international trade." *Journal of International Economics*, 48(1): 7–35.

Romer, Paul M. 1986. "Increasing returns and long-run growth." *The journal of political economy*, 1002–1037.

Rosenblat, Tanya S, and Markus M Mobius. 2004. "Getting closer or drifting apart?" *The Quarterly Journal of Economics*, 971–1009.

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.

Scott, John, and Peter J Carrington. 2011. *The SAGE handbook of social network analysis.* SAGE publications.

Shiller, Robert J. 2007. "Understanding recent trends in house prices and home ownership." National Bureau of Economic Research.

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.

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review*, 95(1): 450–460.

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

Verbrugge, Lois M. 1983. "A Research Note on Adult Friendship Contact: A Dyadic Perspective." *Social Forces*, 62(1): 78–83.

Wikipedia. 2005. "Distribution of Norwegian Americans according to the 2000 census." https://en.wikipedia.org/wiki/File:Norwegian1346.gif, accessed: 2016-12-21.

Yilmazkuday, Hakan. 2012. "Understanding interstate trade patterns." *Journal of International Economics*, 86(1): 158–166.

Zipf, George Kingsley. 1949. "Human behavior and the principle of least effort."

A Appendix

In this Appendix, we further explore the ability of the Social Connectedness Index to help us understand the geography of social networks within the United States.

A.1 Material Related to Main Body of Paper

We first describe in more detail a number of extensions of Tables and Figures referenced in the main body of the paper. First, Appendix Table A1 displays the geographic concentrations of friendship networks. This measure is similar to the one presented in Table 1, but is calculated for only the 48 continental states (i.e., it excludes Hawaii and Alaska). The key facts regarding the geographic concentration of friendship networks from the main body of the paper are unaffected.

Appendix Figure A1 displays the friendships networks of New York County, NY (coterminous with Manhattan), and Bronx County, NY (coterminous with the Bronx), with all other counties in the continental United States, as a companion to the maps in Figure 1. Panels A and B show the share of the friends of residents of Manhattan and the Bronx, respectively, in each county in the continental United States, constructed as in equation 1. The geographic spread of the friendship networks of Manhattan and the Bronx look similar in this specification, with Manhattan having a higher share of friends in most of areas away from the East Coast. Panels C and D are constructed as in equation 2, showing the relative probability of a connection to each county in the continental United States. In this specification, Manhattan's social network shows a much wider spread across the Midwest, the Mountain States, and the West Coast, while the Bronx shows more connectivity to a number of counties in Southern states like Louisiana, Mississippi, and Alabama.

Appendix Figures A2 and A3 present county-level binned scatter plots of the share of friends living within 100 miles by demographic characteristics, as in Figure 6, but conditioning on state and commuting zone, respectively. Most of these plots show patterns very similar to those in Figure 6. 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 the figure in the main text, the plot shows social mobility to decline as the share of friends within 100 miles increases; including fixed effects for state and commuting zone causes this relationship to reverse, with social mobility rising as the share of friends within 100 miles increases.

Appendix Figures A4, A5, and A6 present county-level binned scatter plots of the share of friends living among the nearest 50 million people by demographic characteristics. Figures A5 and A6 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 A4, A5, and A6 than the corresponding Panels in Figure 6 and Appendix Figures A2 and A3. The R^2 of the quadratic regressions underlying each Panel in Appendix Figure A4 are included in the main text.

We also conduct a multi-variate analysis between our measures of geographic concentration and socioeconomic outcomes at the county level. In columns 1 to 3 of Appendix Table A2, 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. 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.

Appendix Figure A7 shows binned scatter plots at the county-pair level that portray the univariate relationship between differences across the two counties along a number of outcome variables and the social connectedness between them. 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.

A.2 Additional Exploration of Social Connectedness

The rest of the Appendix discusses a number of additional interesting patterns observed in the SCI data. These patterns further highlight the ability of these data to provide important insights into the geography of U.S. social networks. In all Figures focusing on within-U.S. connectedness, we plot the scaled relative probability that a given user in county j has a friendship link to a given user in county i, as constructed by equation 2 in the main body of the paper.

Effect of State and Regional Borders. We begin by further exploring the role of state and regional borders in shaping social connectedness. In particular, Table 2 and Figure 4 in the main text already demonstrated 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 A8. In each of the panels, the friendship networks are most dense within the state, 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 A8 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 Panel B of Figure 4, 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 Figure 2). Panel G of Appendix Figure A8 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 perform regression A1 separately for each county i. The unit of observation is a county-pair. The dependent variable, $\log(f_{ij})$, denotes the log of the number of friendship links between counties i and j (i.e., 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.

$$\log(f_{ij}) = \beta_0 + \beta_1 * \mathbb{1}_{\text{Same State}} + \beta_2 g(d_{ij}) + \epsilon_{ij} \tag{A1}$$

Appendix Figure A9 maps the coefficient β_1 for each county in the continental United States. Red counties reflect 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 A10 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 A11, A12, and A13 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 6 and Appendix Figures A2 and A3. 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 A14 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 Panels B and C of Figure 4, 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-U.S. Social Connectedness. While we have previously documented a number of strong patterns in social connectedness across U.S. counties (i.e., it declines in geographic distance, and state borders and physical barriers matter), 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 A15 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 New York (Appendix Figure A1 shows that connections to the South are present for both the Bronx and Manhattan), Chicago, and, as discussed in the following paragraph, Milwaukee. Panel C of Figure A15 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 A19, and the associated discussion).

Counties within Wisconsin also display significant heterogeneity in the geographic distribution of their friendship networks. Panel A of Appendix Figure A16 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 (see Section 2).

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 A17 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 A18 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 A19 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 A20 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.

Further Explorations of International Social Connectedness. We also explore a number of additional dimensions of social connectedness of U.S. counties to foreign countries. In particular, Appendix Figure A21 shows the share of friendship links in each county in the continental United States to various countries (see also Figure 13). Panel A 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 B 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 C 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 D and E 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 F 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 Bibliography

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

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

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

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.

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.

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.

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

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.

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, Contiguous 48

	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	55.5%	62.9%	70.4%	79.9%	1.3%	2.8%	6.7%	22.6%
P5	38.4%	46.5%	55.2%	64.5%	0.1%	0.3%	1.2%	5.7%
P10	42.5%	49.6%	57.4%	67.4%	0.1%	0.6%	2.1%	8.2%
P25	48.6%	56.0%	63.9%	75.1%	0.3%	1.1%	3.6%	14.1%
Median	55.5%	63.9%	71.8%	82.0%	0.8%	2.2%	5.9%	22.7%
P75	63.2%	70.9%	78.0%	86.3%	1.8%	3.5%	8.3%	30.9%
P90	67.4%	74.8%	81.2%	89.0%	3.3%	6.3%	15.1%	37.3%
P95	70.3%	77.0%	83.2%	91.0%	5.4%	9.2%	15.8%	40.1%

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 U.S. population living within a certain distance. Counties are weighted by their population. Shares are calculated for the contiguous 48 states only.

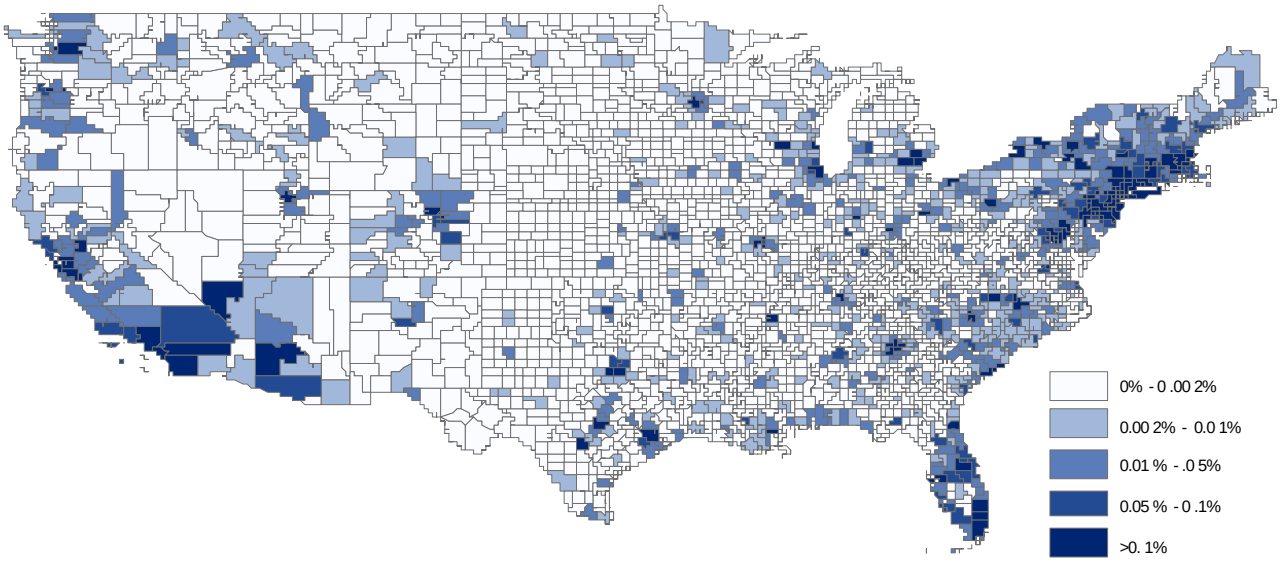
Table A2: Concentration of Networks and Socioeconomic Outcomes

	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

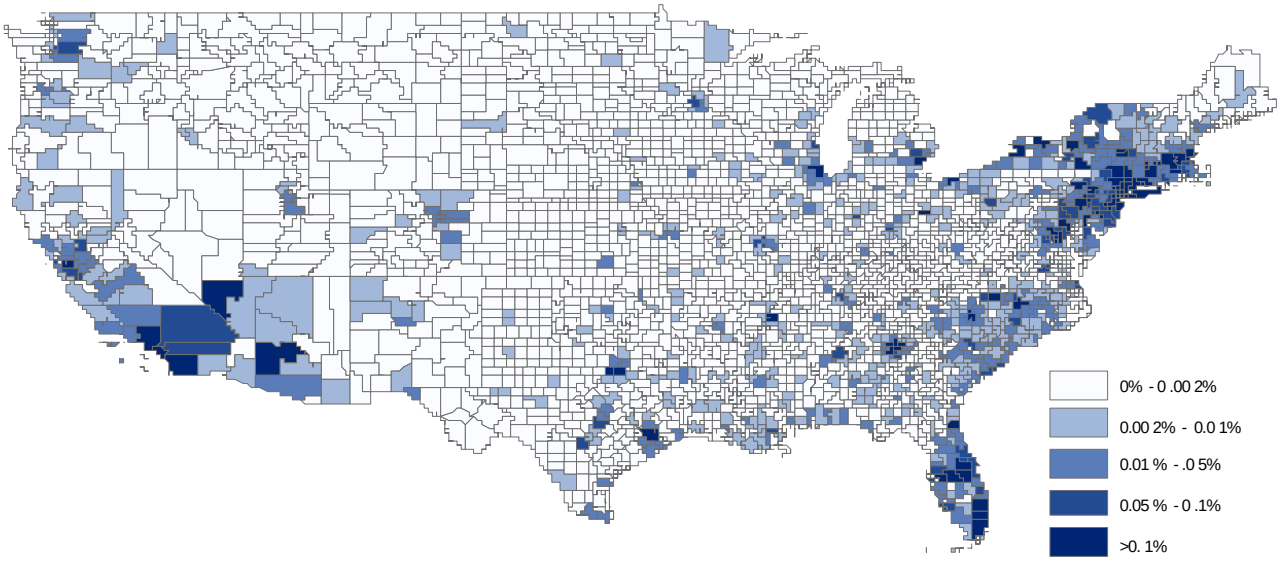
Note: Table shows results from a regression of county-level measures of the concentration of social networks on measure of county-level socioeconomic outcomes. In columns 1 to 3, we analyze the correlation with the share of friends living within 100 miles, in columns 4 to 6 the share of friends within 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.

Figure A1: County-Level Friendship Maps, New York

(A) New York County (Manhattan), NY - Share of Friendship Links ($ShareFriends_{i,j}$)



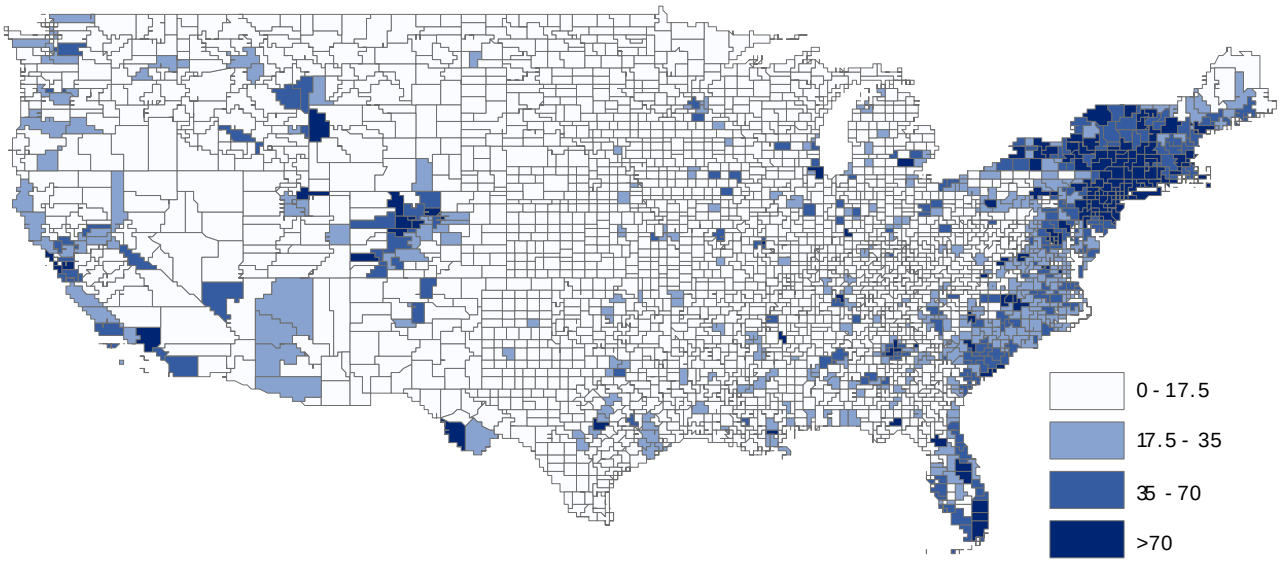
(B) Bronx County (The Bronx), NY - Share of Friendship Links ($ShareFriends_{i,j}$)



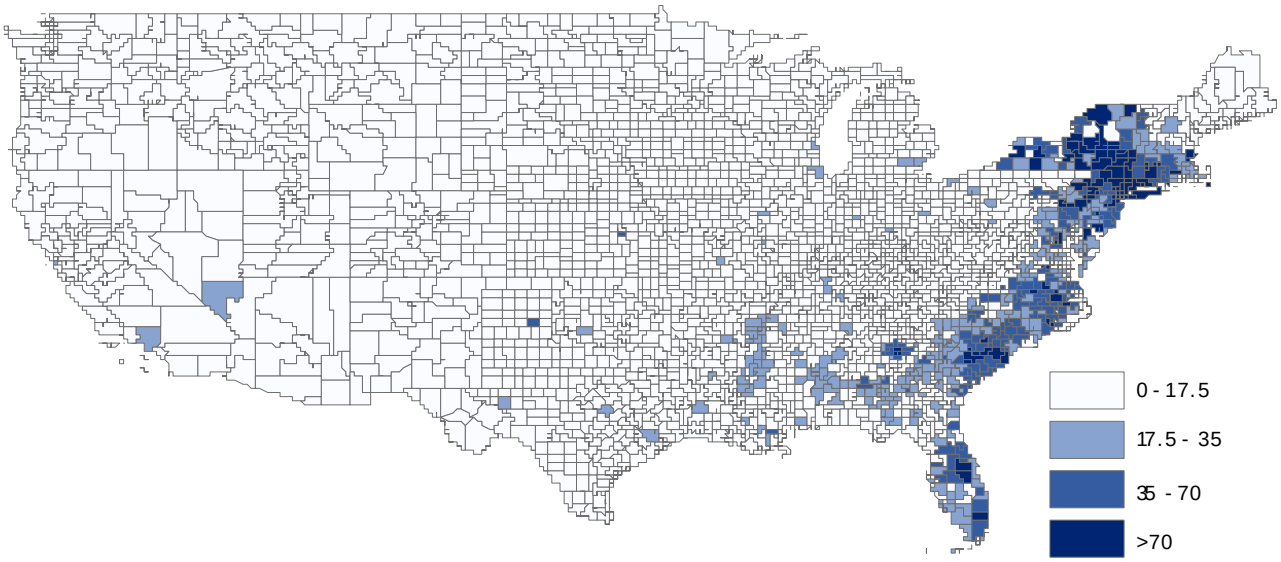
Note: Figure shows the share of friendship links of New York County, NY (coterminous with the New York City borough of Manhattan) (Panel A) and Bronx County, NY (coterminous with New York City borough of the Bronx) (Panel B) to all other counties in the continental United States, constructed as in equation 1. Darker colors correspond to counties in which the home-county's Facebook users have a larger share of friends.

Figure A1: County-Level Friendship Maps, New York

(C) Relative Probability of Friendship Link to New York County (Manhattan), NY ($RelativeProbFriendship_{i,j}$)

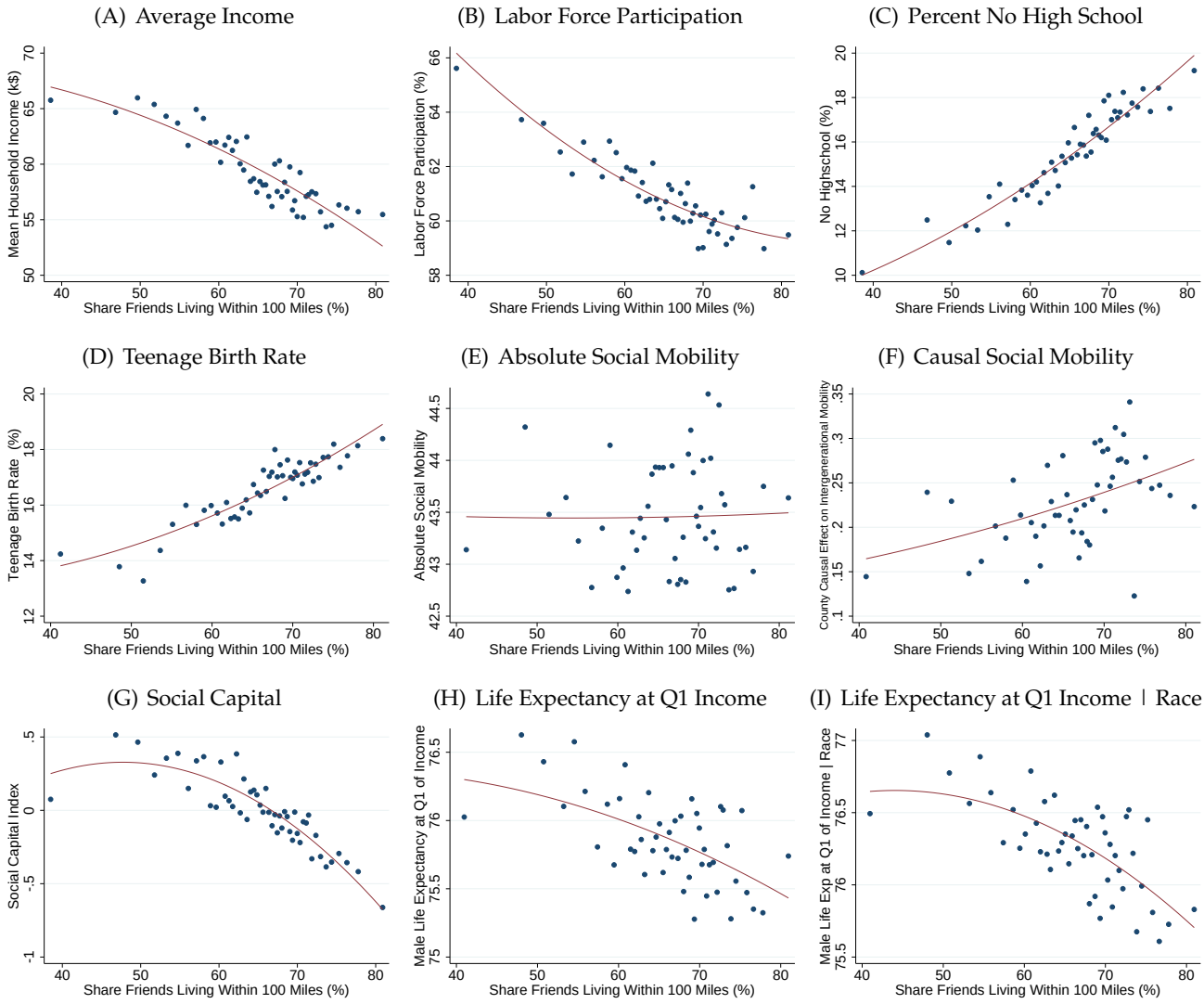


(D) Relative Probability of Friendship Link to Bronx County (The Bronx), NY ($RelativeProbFriendship_{i,j}$)



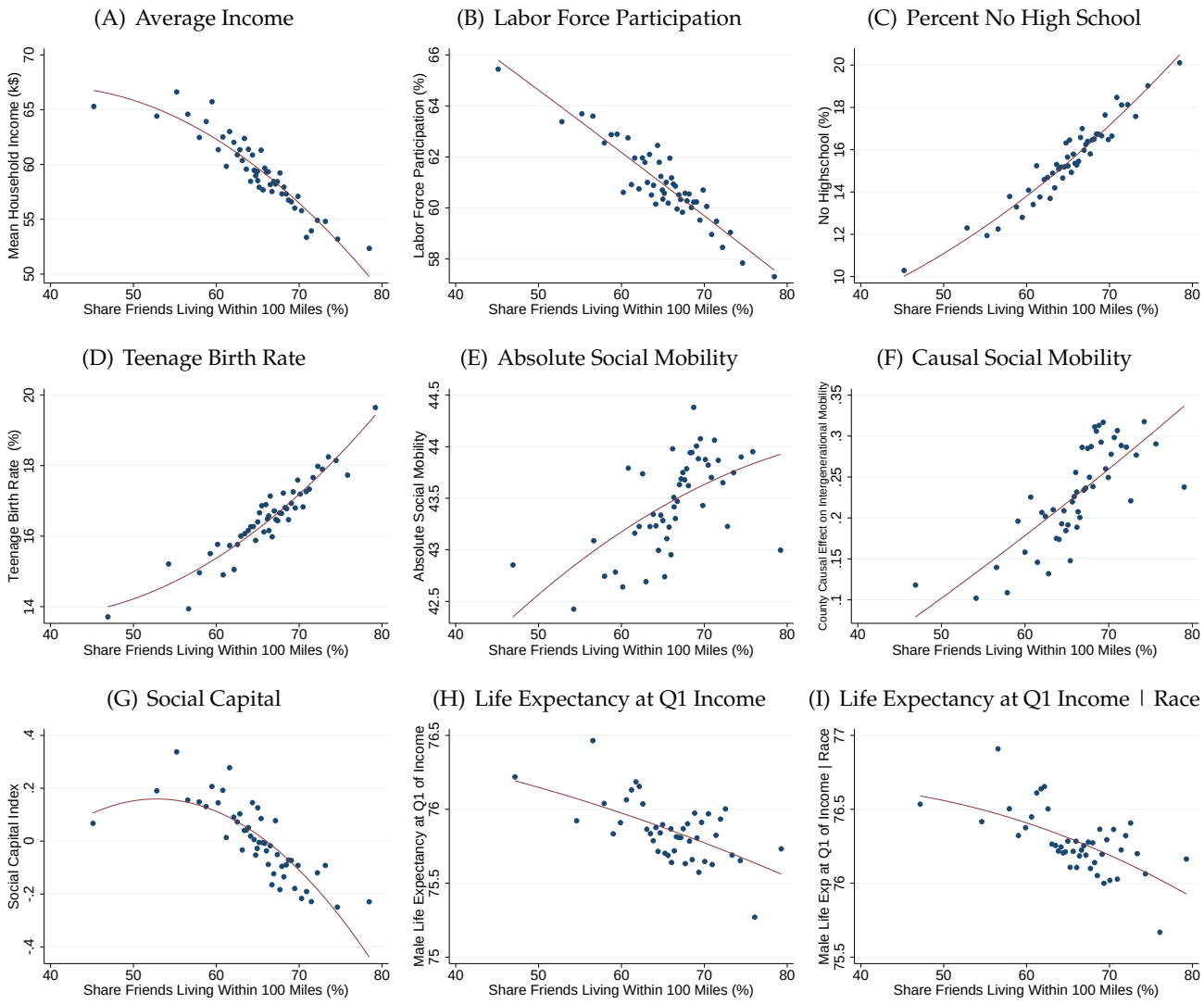
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to New York County, NY (coterminous with the New York City borough of Manhattan) in Panel C, and Bronx County, NY (coterminous with New York City borough of the Bronx) in Panel D. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (New York or Bronx) and county j.

Figure A2: Share of Friends Within 100 Miles - Conditional on State



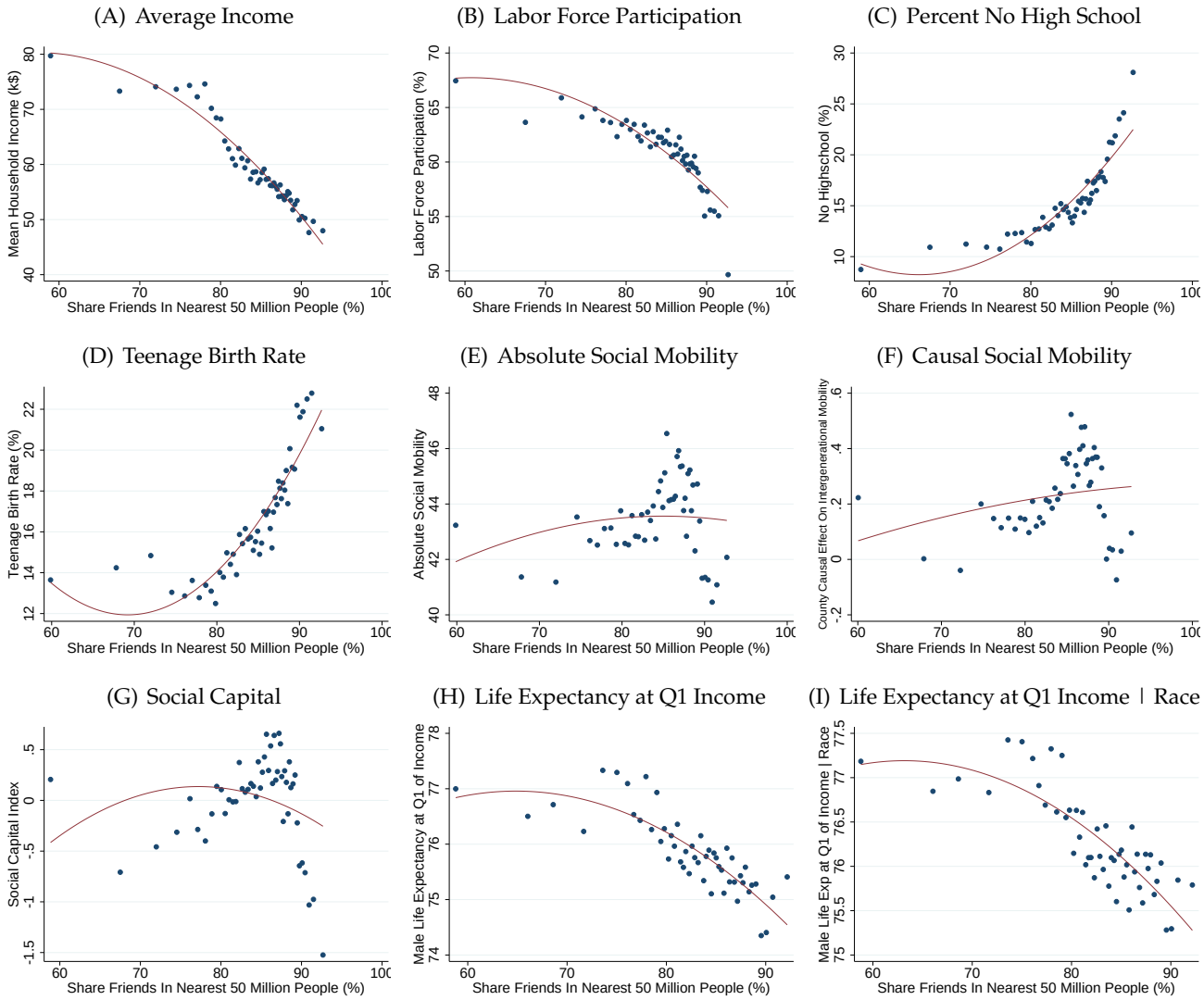
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force 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 A3: Share of Friends Within 100 Miles - Conditional on Commuting Zone



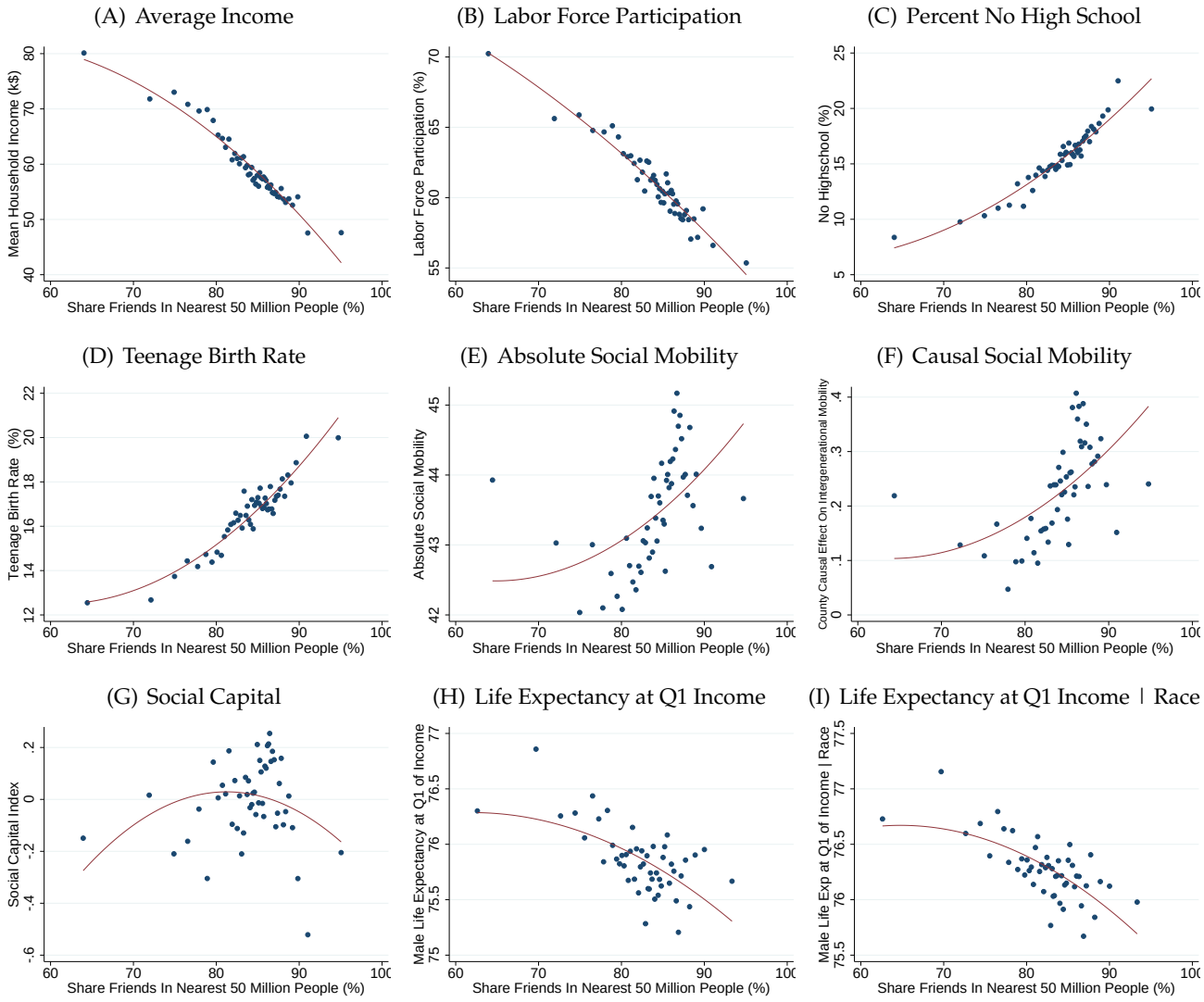
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the county's labor force 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 A4: Share of Friends Among Nearest 50 Million People



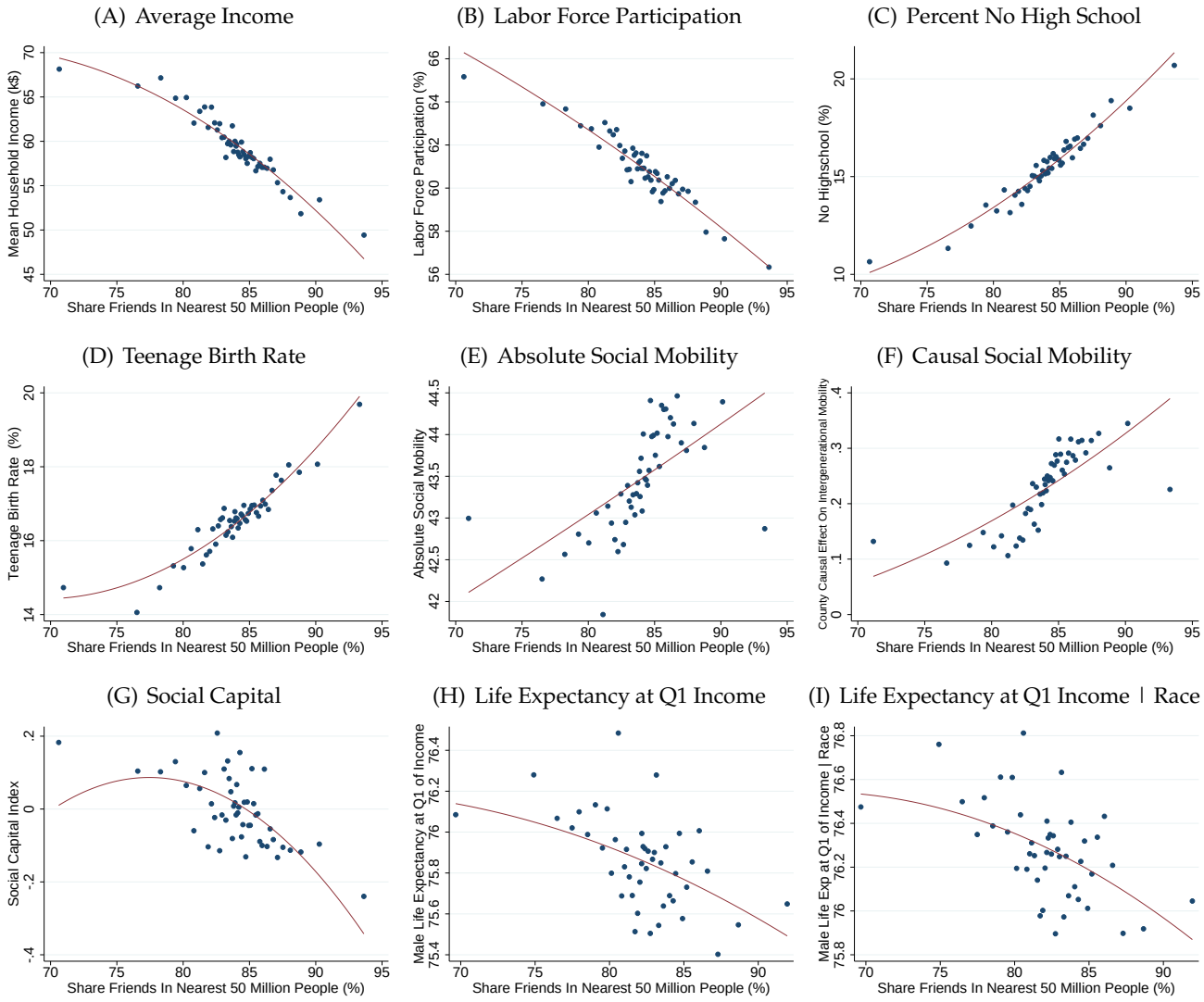
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within the nearest 50 million people. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the 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 A5: Share of Friends Among Nearest 50 Million People - Conditional on State



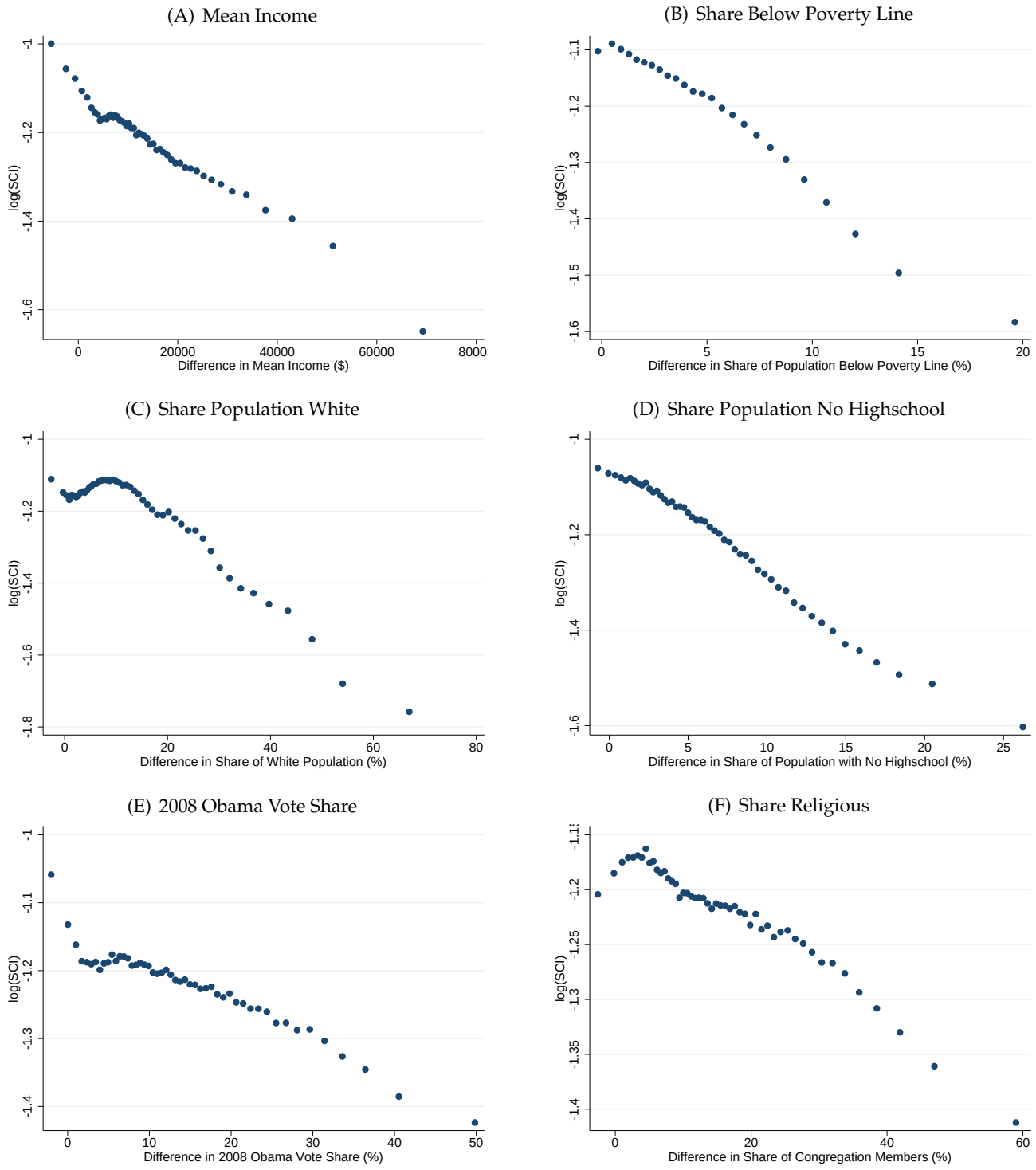
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure the share of friends that live within the nearest 50 million people. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income (Panel A), the 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 A6: 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.

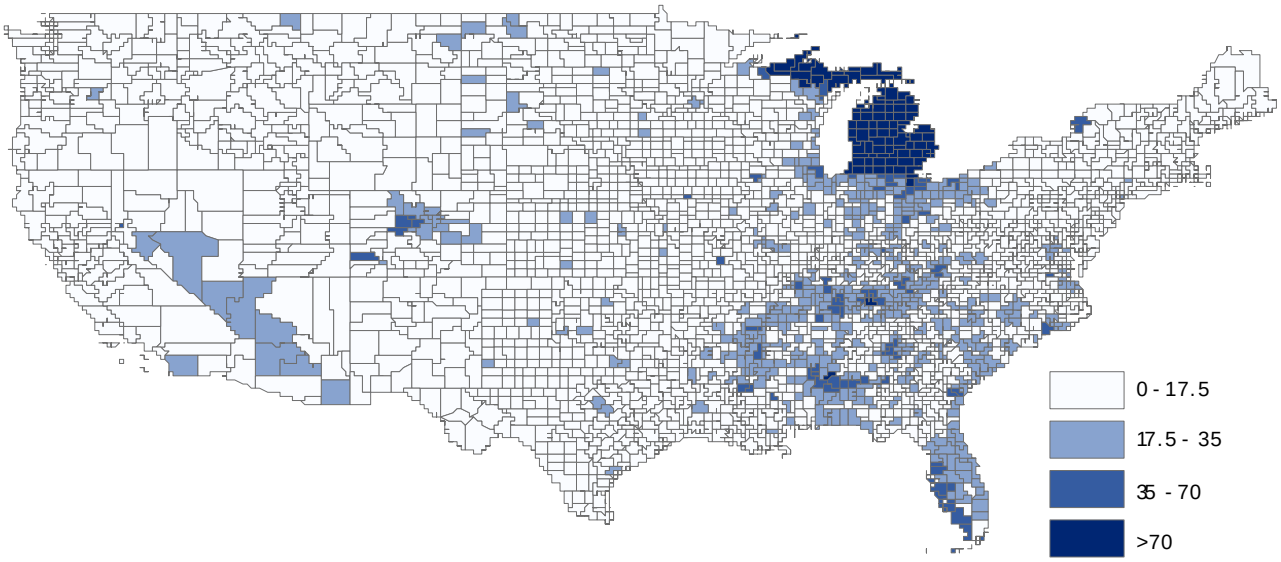
Figure A7: The Geographic Spread of Friendship Networks and County-Level Outcomes



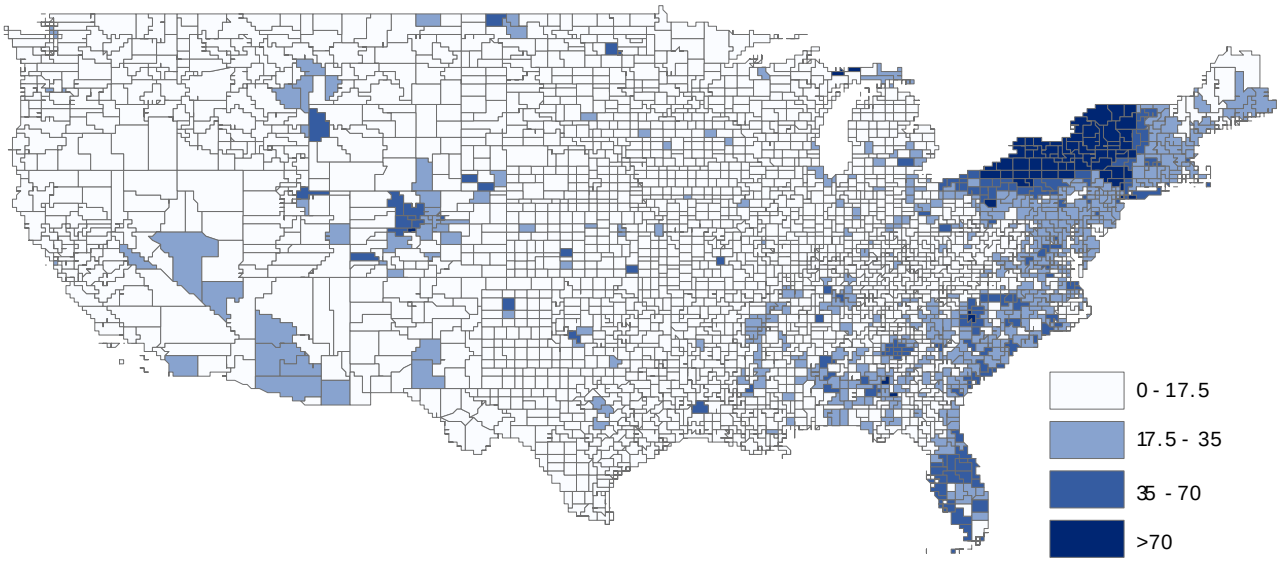
Note: Figure shows binned scatter plots, with county-pairs as the unit of observation. On the horizontal axis is the difference across the county-pairs for a number of county-level measures: mean income (Panel A), share of population below poverty line (Panel B), share of population that is white (Panel 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 A8: State Borders and Regional Groupings

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



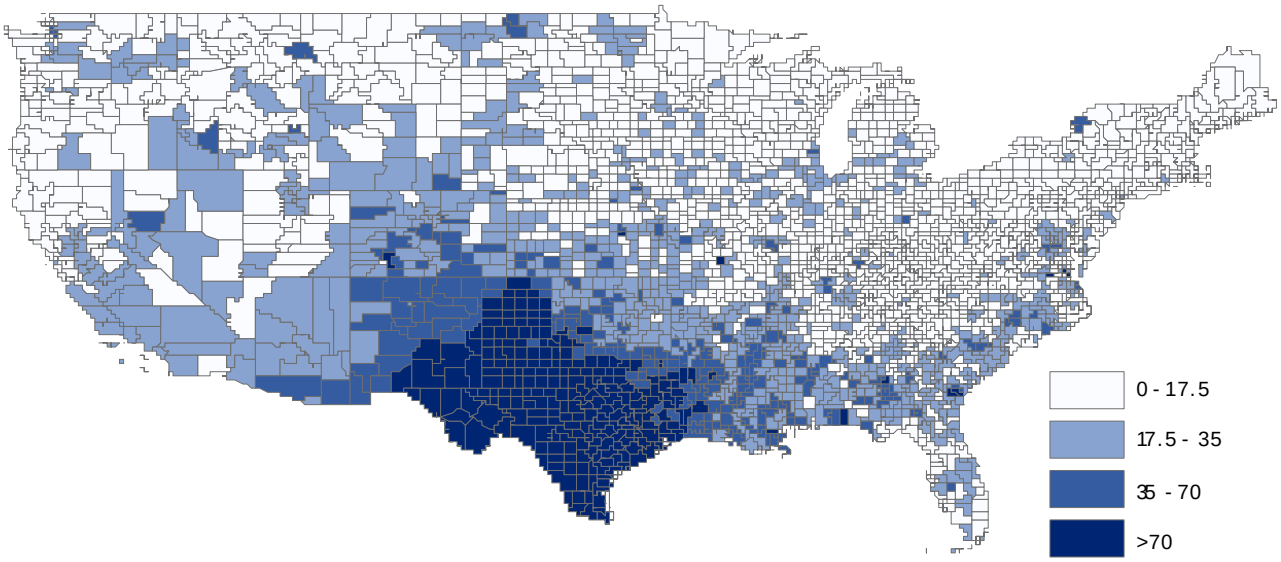
(B) Relative Probability of Friendship Link to Erie County, NY ($RelativeProbFriendship_{i,j}$)



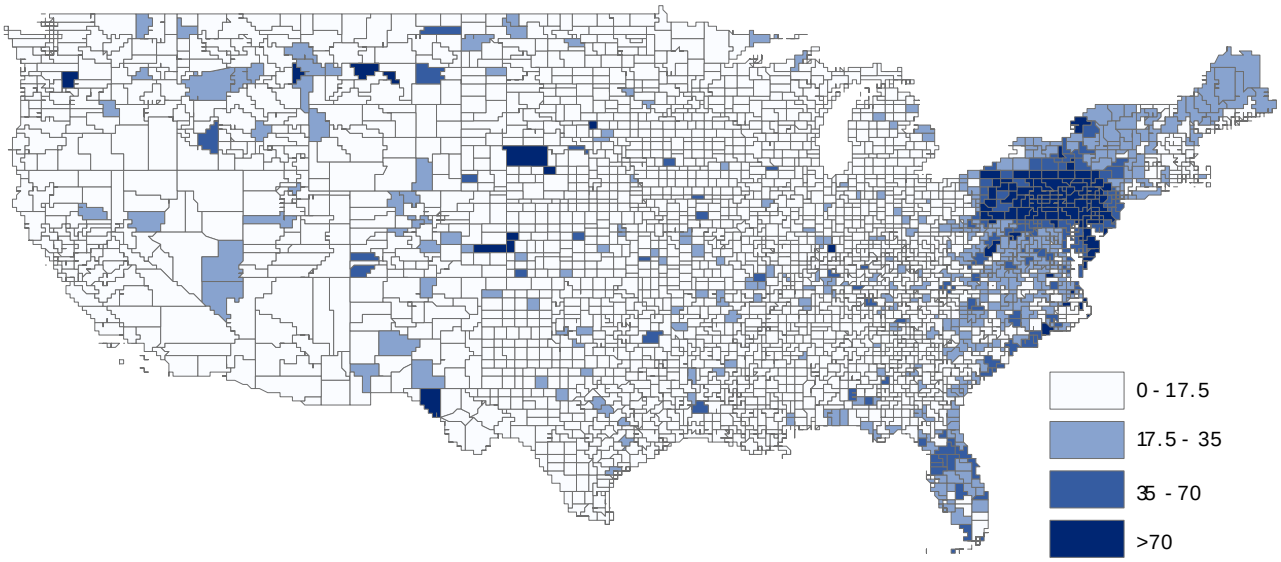
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Macomb County, MI in Panel A, and Erie County, NY in Panel B. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Macomb or Erie) and county j.

Figure A8: State Borders and Regional Groupings

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



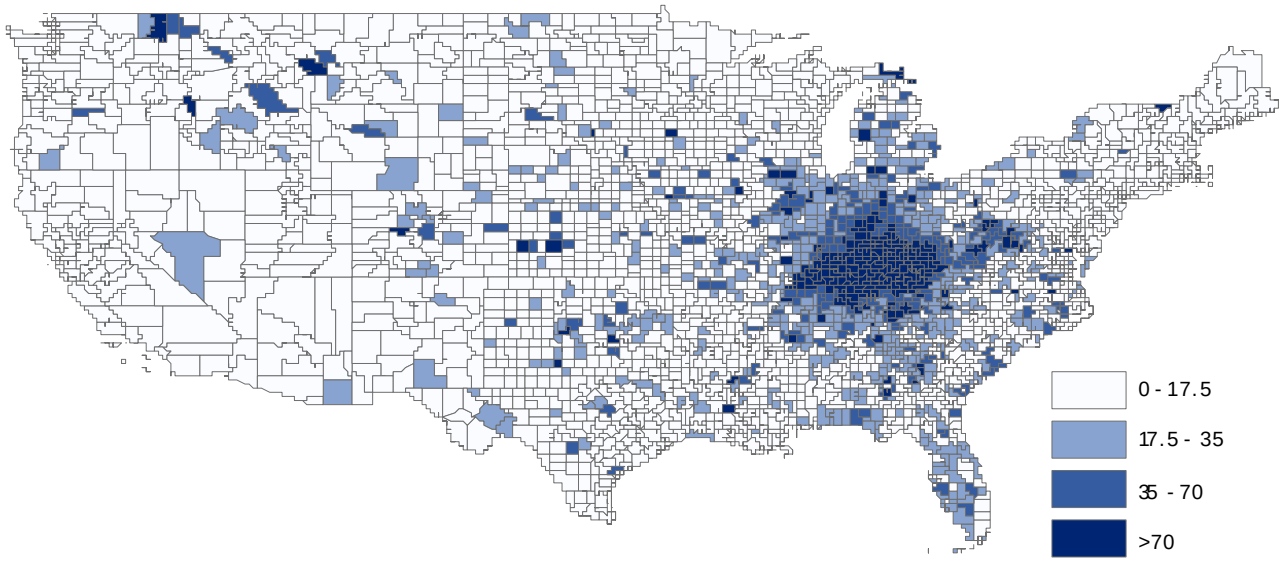
(D) Relative Probability of Friendship Link to Schuylkill County, PA ($RelativeProbFriendship_{i,j}$)



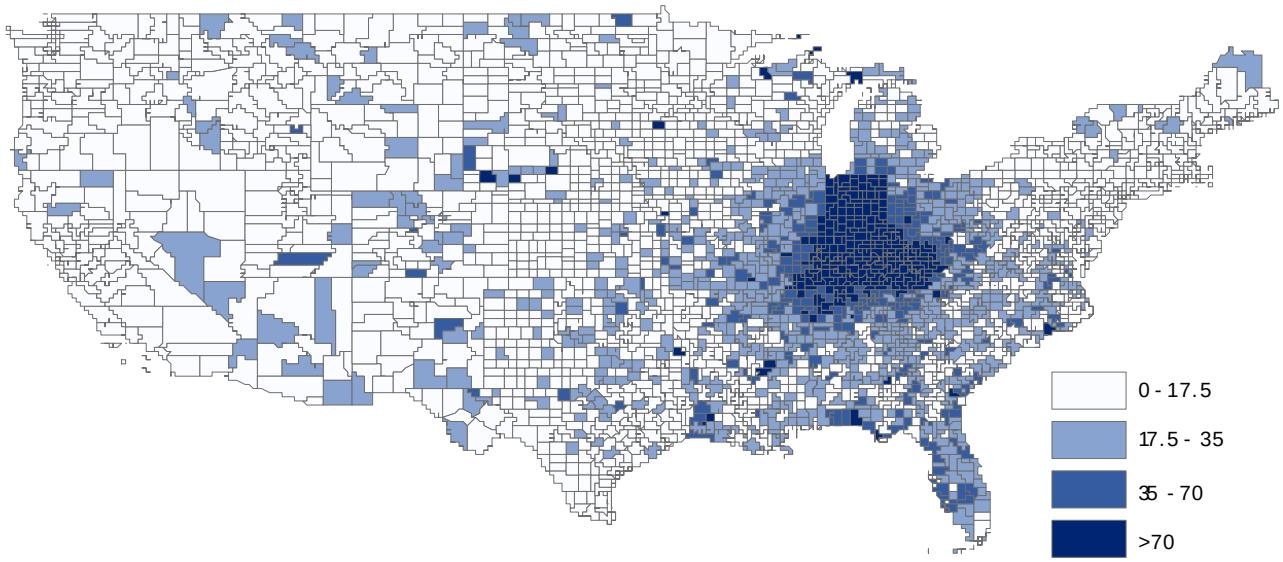
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Bexar County, TX in Panel C, and Schuylkill County, PA in Panel D. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Bexar or Schuylkill) and county j.

Figure A8: State Borders and Regional Groupings

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



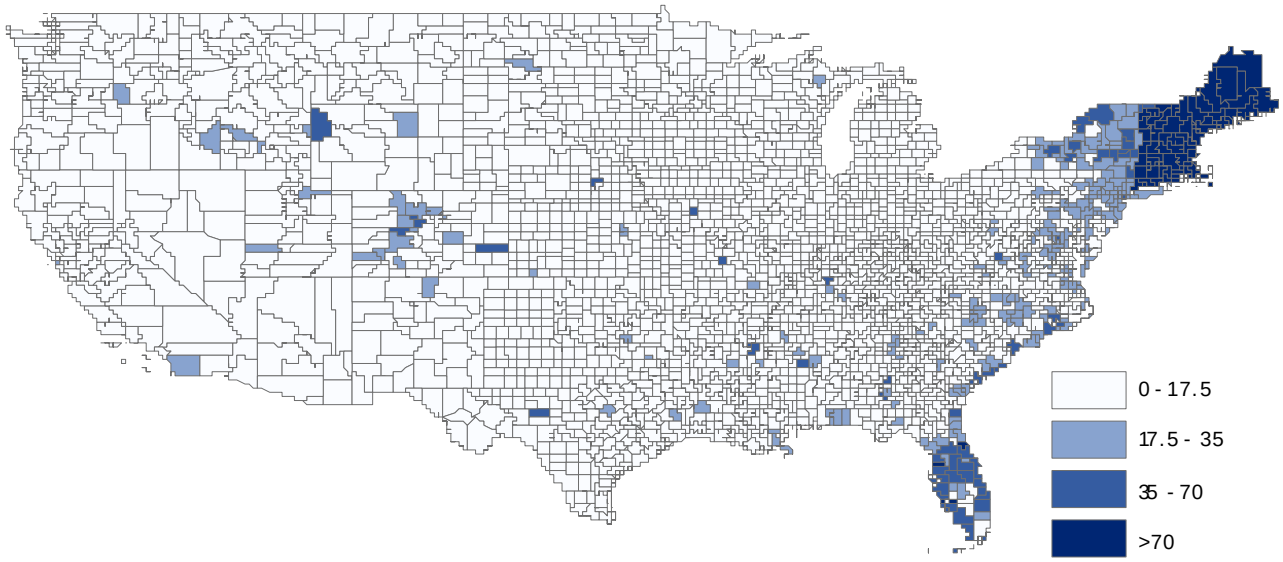
(F) Relative Probability of Friendship Link to Clark County, IN ($RelativeProbFriendship_{i,j}$)



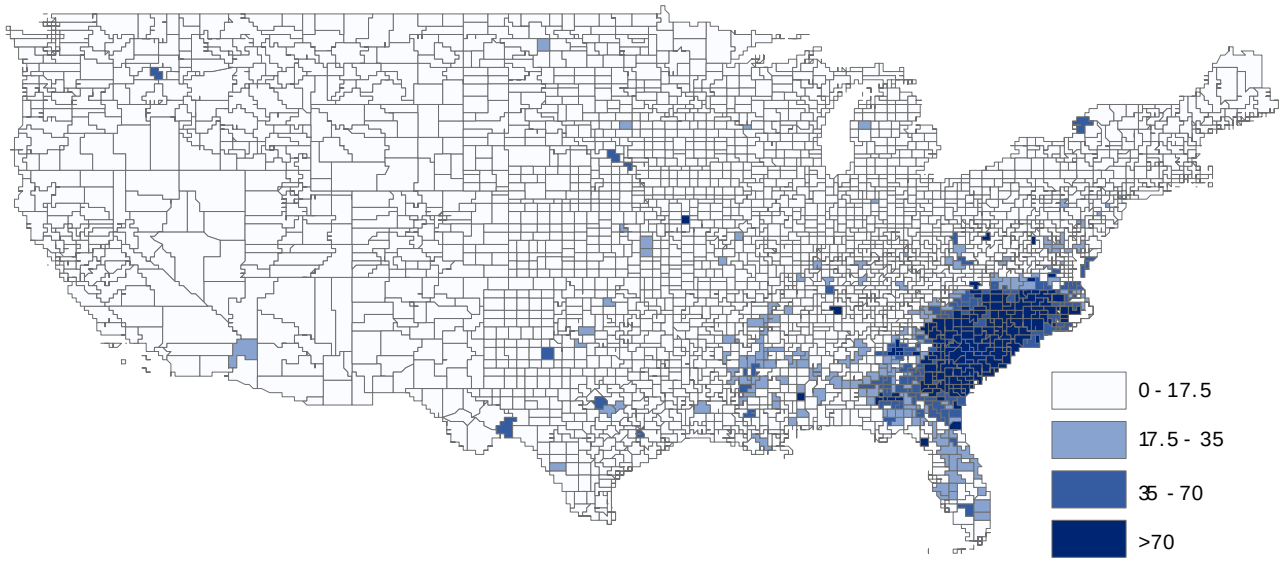
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Marion County, KY in Panel E, and Clark County, IN in Panel F. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Marion or Clark) and county j.

Figure A8: State Borders and Regional Groupings

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

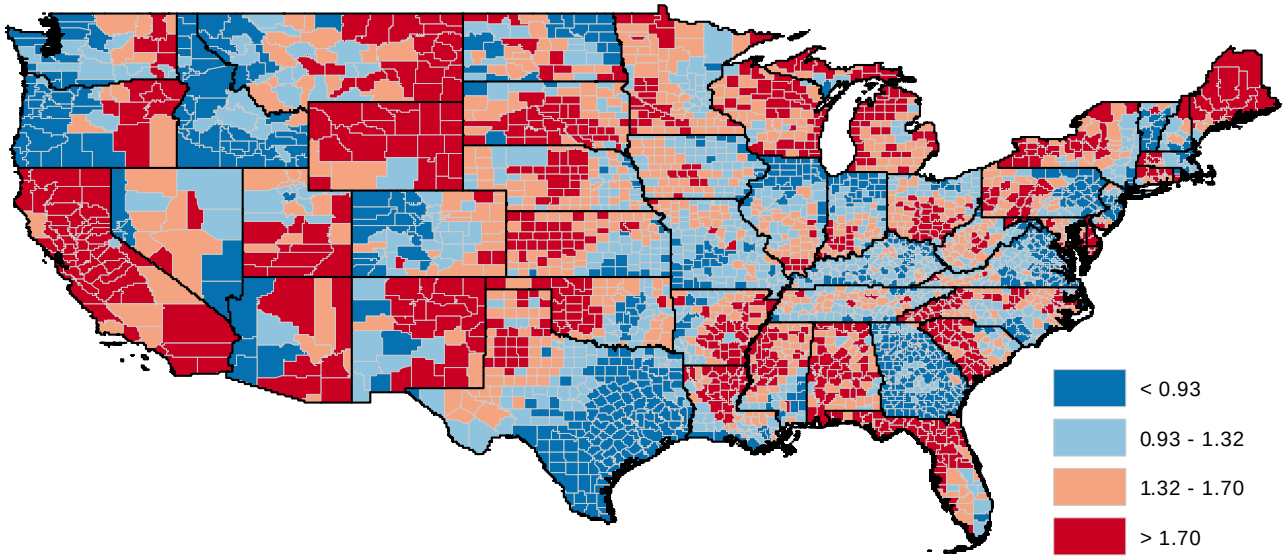


(H) Relative Probability of Friendship Link to Allendale County, SC ($RelativeProbFriendship_{i,j}$)



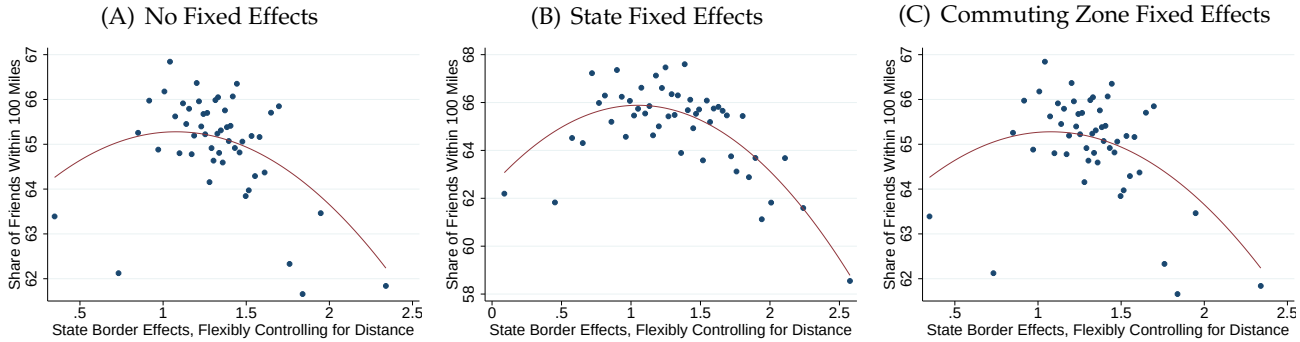
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Bristol County, MA in Panel G, and Allendale County, SC in Panel H. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Bristol or Allendale) and county j.

Figure A9: Magnitude of State Border Effects



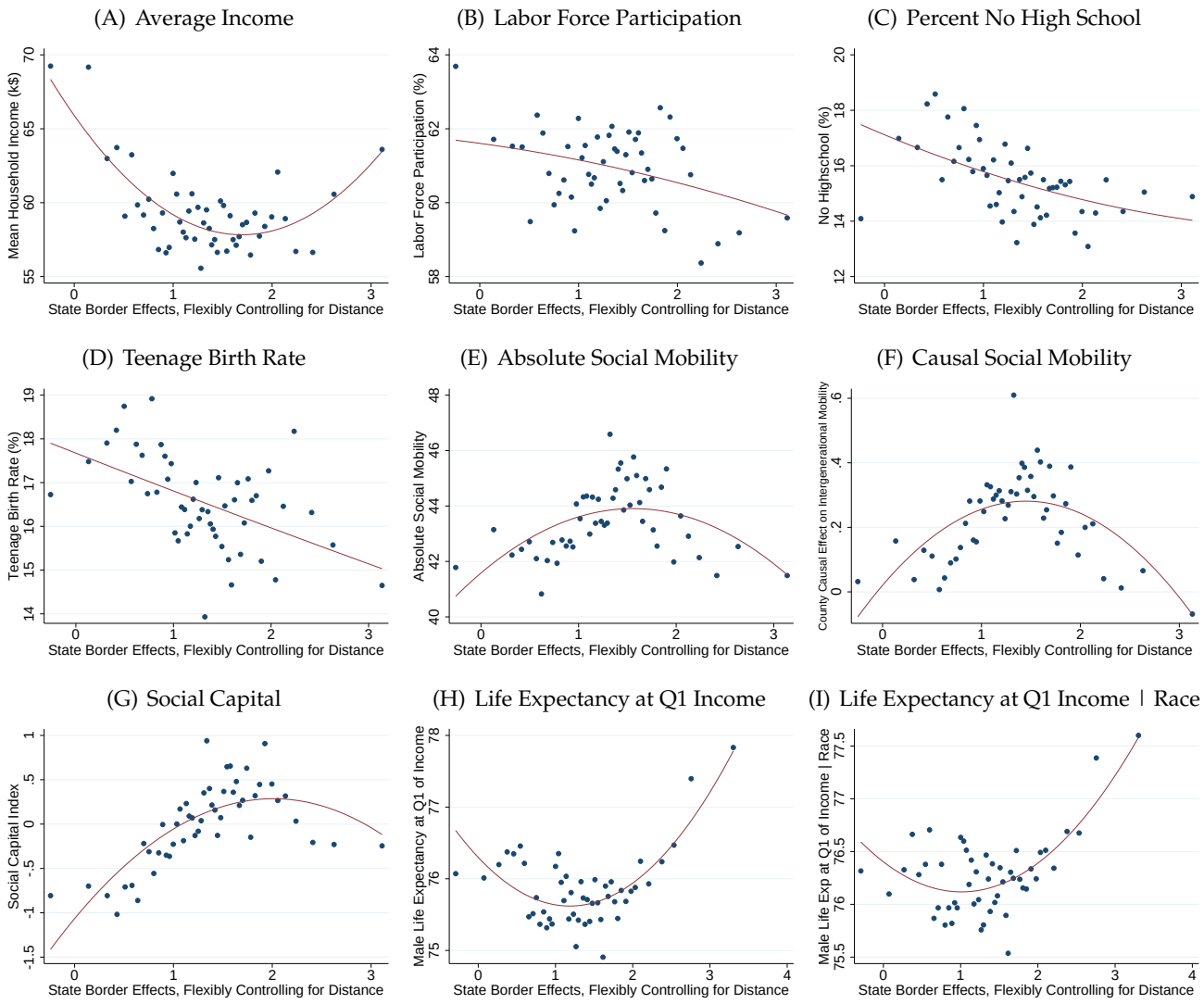
Note: This map plots the estimated state border effects by county, given as the estimated value of coefficient β_1 in Regression A1. Counties with a stronger state-border effect are red, while counties with a weaker effect are in blue.

Figure A10: State Border Effects and Share of Friends Within 100 Miles



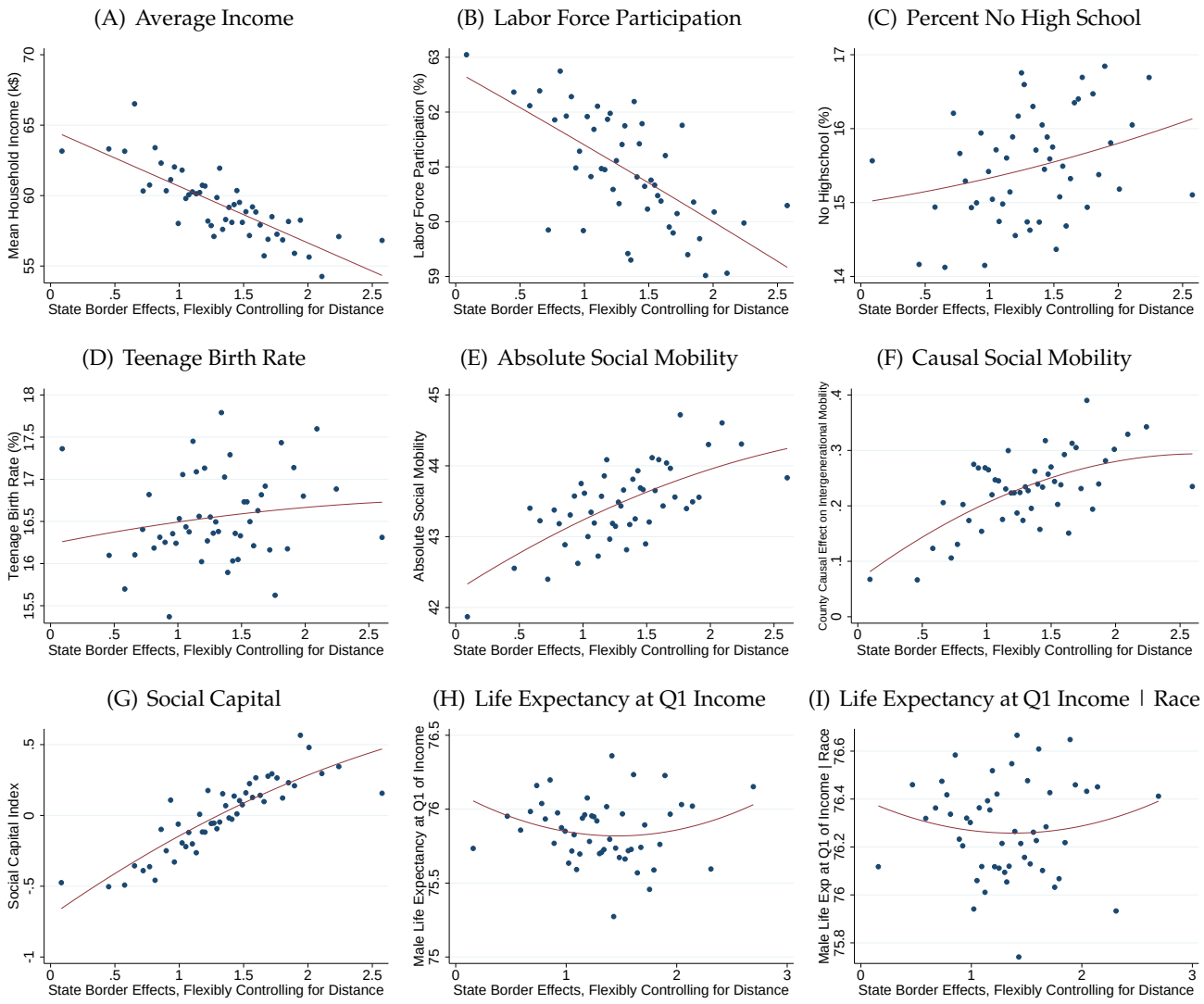
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A1. The vertical axis of each panel shows the share of friends within 100 miles for each bin. Panel A does not include fixed effects, Panel B includes state fixed effects, and Panel C includes commuting zone fixed effects.

Figure A11: State Border Effects Coefficients



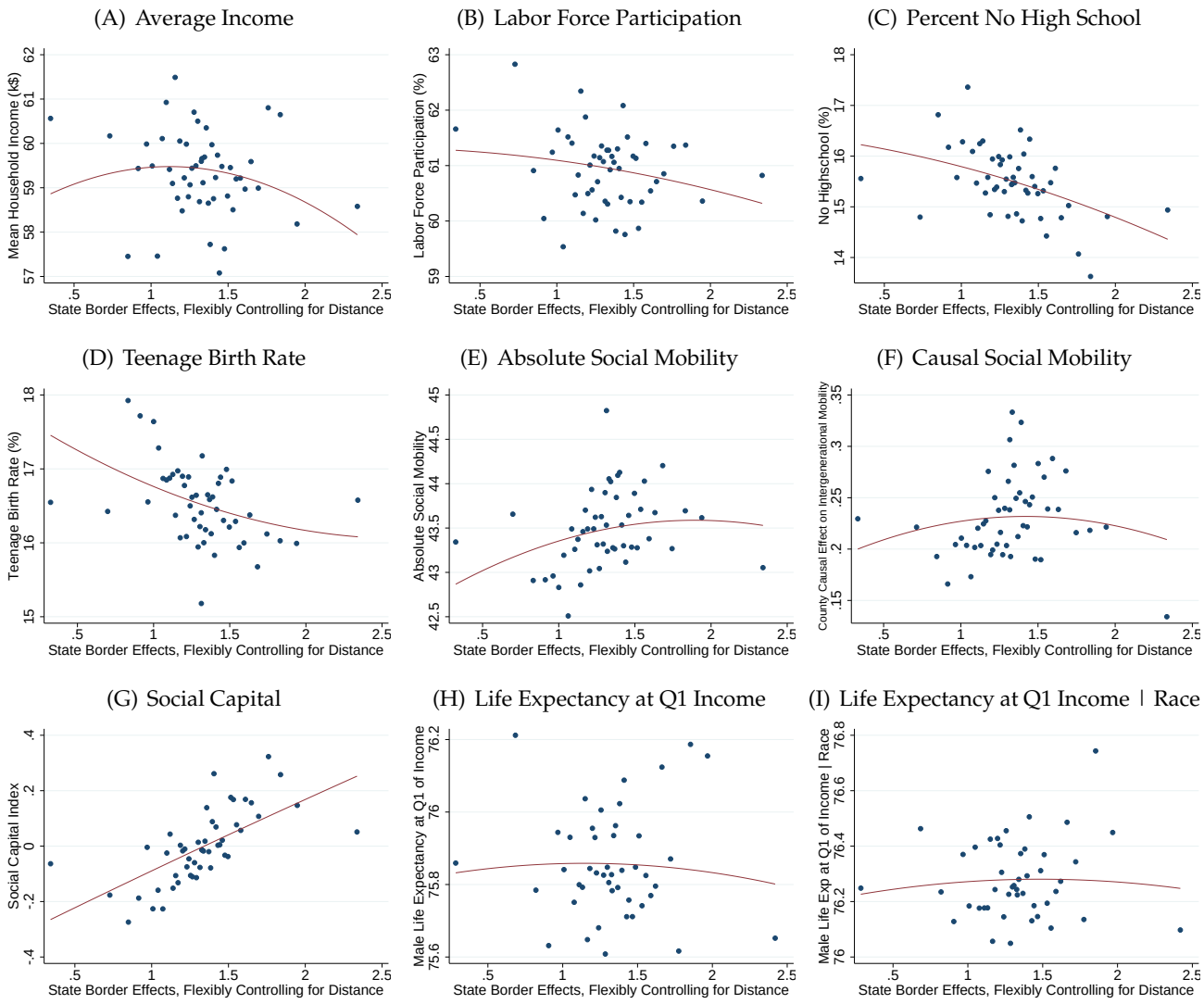
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A1. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the 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 A12: 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 A1. 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.

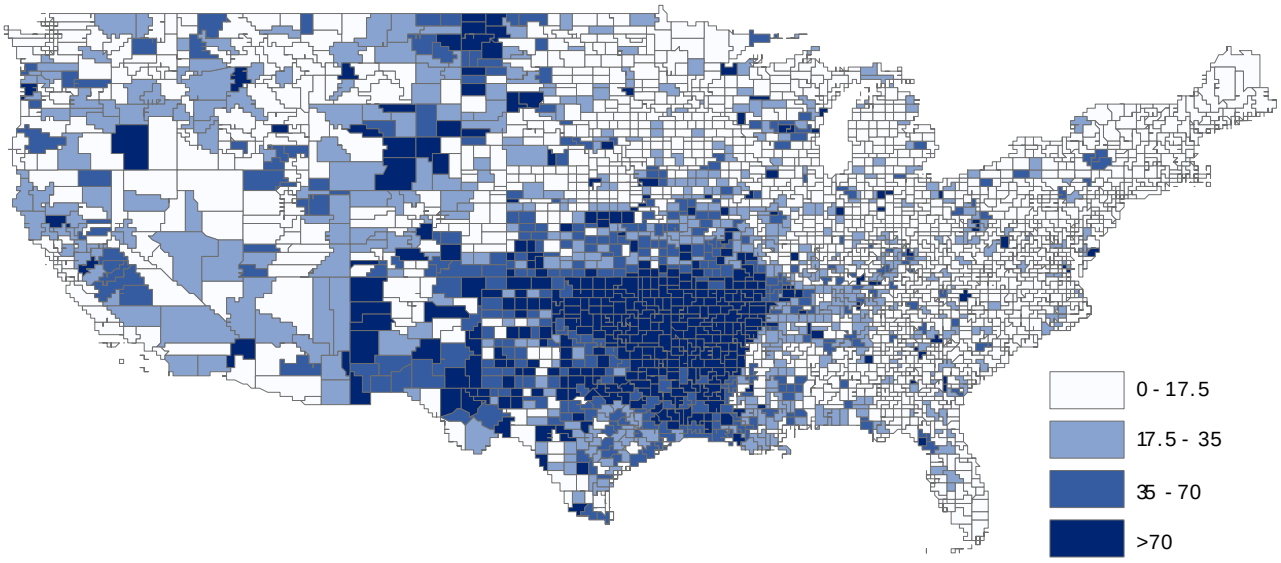
Figure A13: State Border Effects Coefficients - Conditional on Commuting Zone



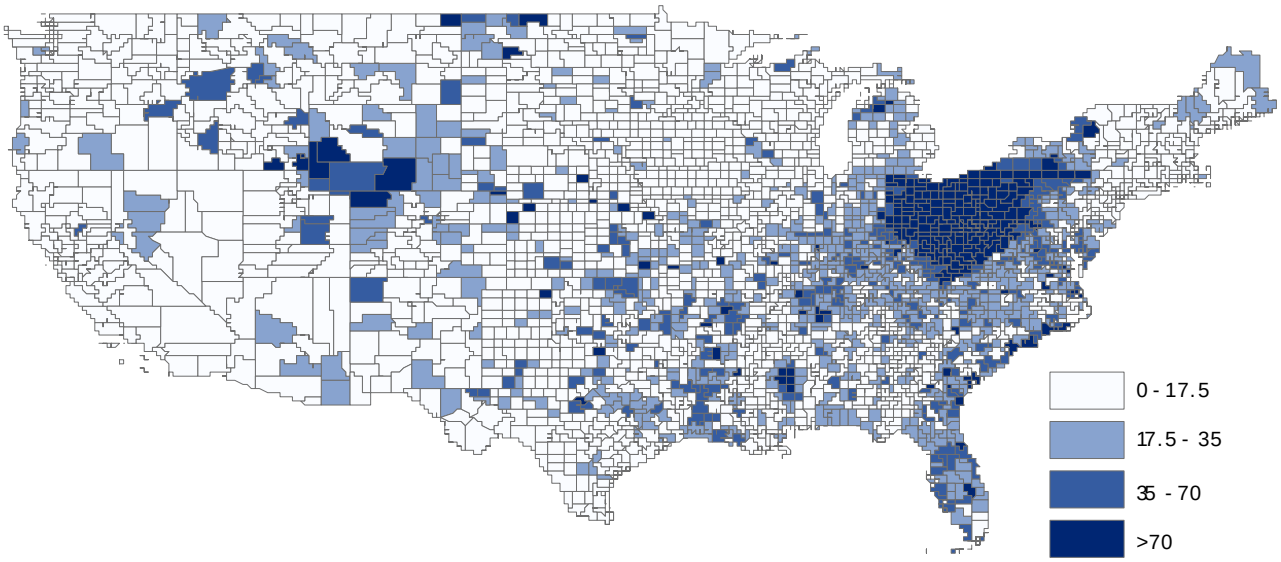
Note: Panels show binned scatter plots, with counties as the unit of observation. The horizontal axes measure estimated state border effects by county, estimated as the value of coefficient β_1 in Regression A1. The regression also controls for commuting zone fixed effects. On the vertical axes are a 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 A14: Geography's Influence on Friendship Network Distribution

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



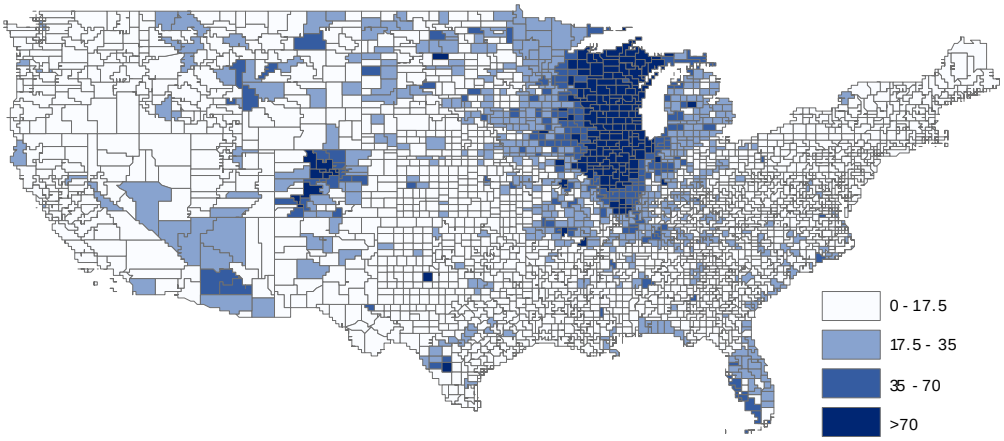
(B) Relative Probability of Friendship Link to Belmont County, OH ($RelativeProbFriendship_{i,j}$)



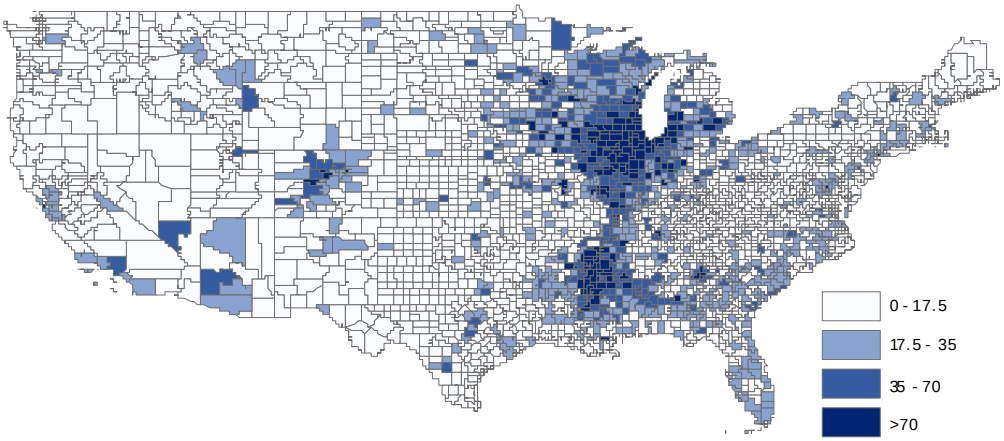
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Scott County, AR, in Panel A, and Belmont County, OH, in Panel B. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Scott or Belmont) and county j.

Figure A15: Heterogeneity in Illinois Friendship Network Distribution

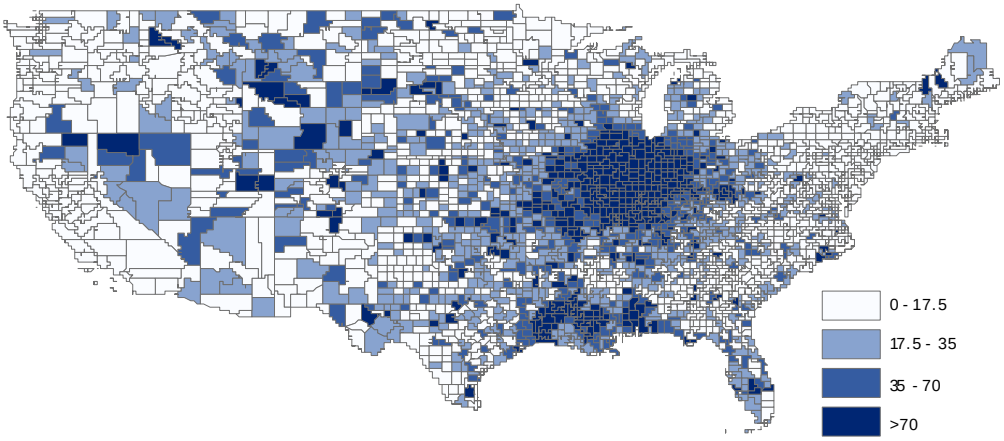
(A) Relative Probability of Friendship Link to McHenry County, IL ($RelativeProbFriendship_{i,j}$)



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



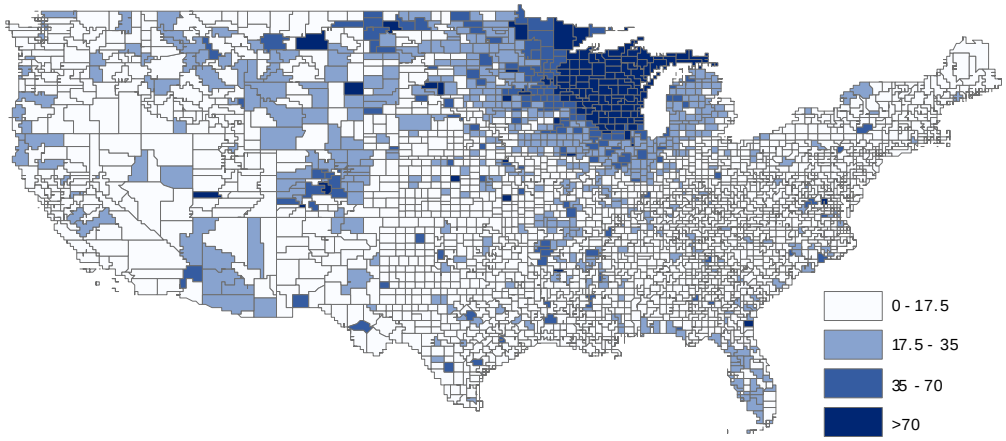
(C) Relative Probability of Friendship Link to Crawford County, IL ($RelativeProbFriendship_{i,j}$)



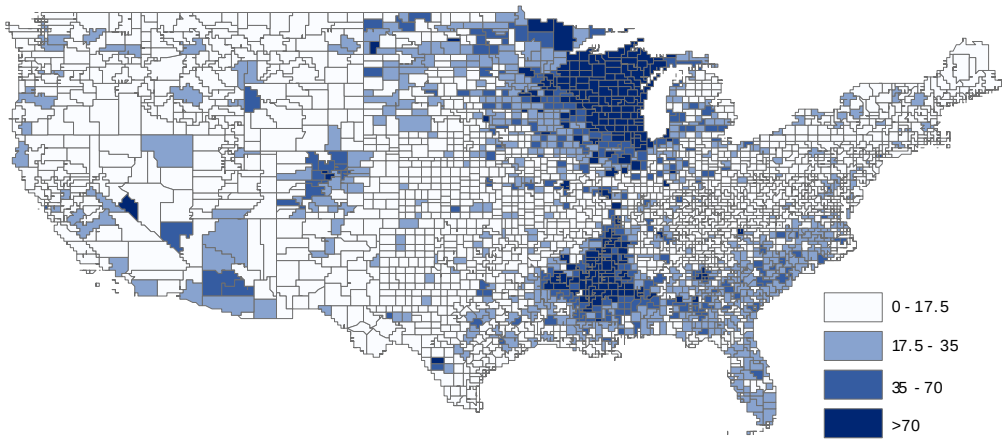
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to McHenry County, IL in Panel A, Cook County, IL in Panel B, and Crawford County, IL in Panel C. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (McHenry, Cook, or Crawford) and county j.

Figure A16: Heterogeneity in Wisconsin Friendship Network Distribution

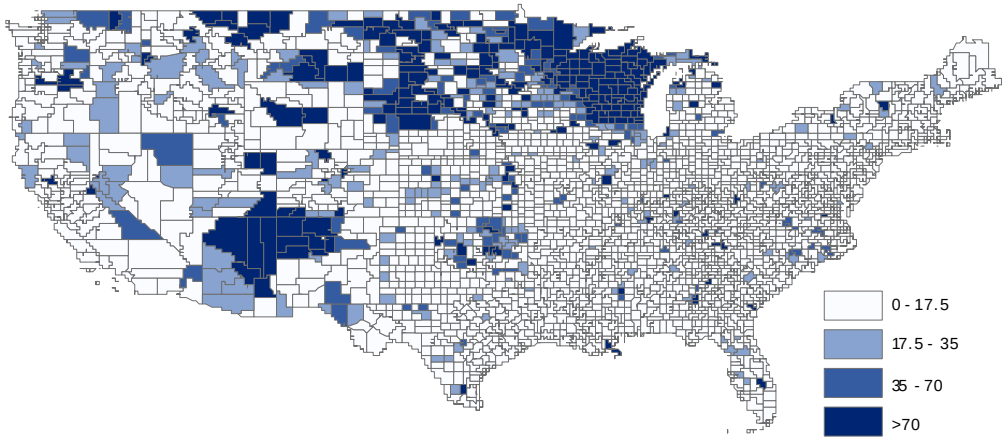
(A) Relative Probability of Friendship Link to Manitowoc County, WI ($RelativeProbFriendship_{i,j}$)



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



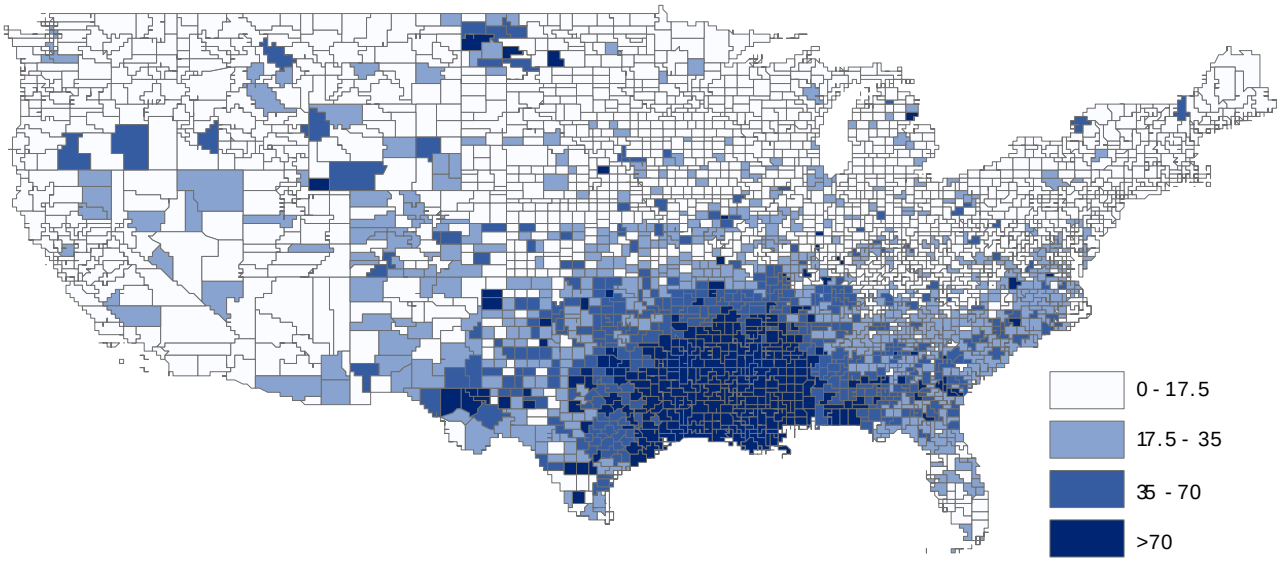
(C) Relative Probability of Friendship Link to Menominee County, WI ($RelativeProbFriendship_{i,j}$)



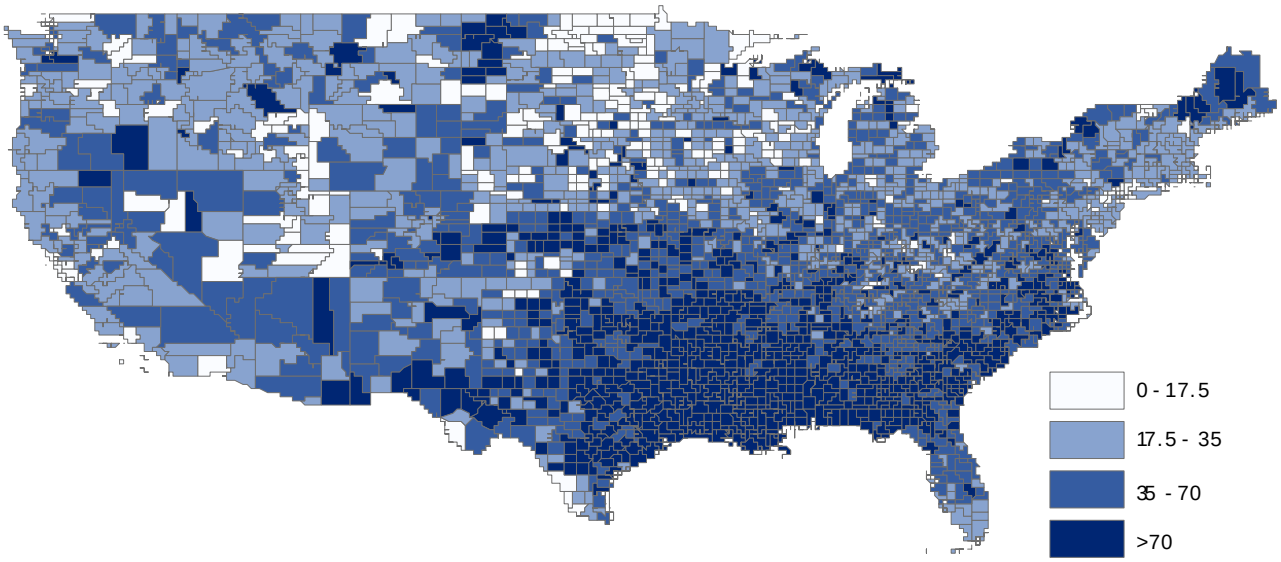
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Manitowoc County, WI in Panel A, Milwaukee County, WI in Panel B, and Menominee County, WI in Panel C. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Manitowoc, Milwaukee, or Menominee) and county j.

Figure A17: Influence of a Military Base on Friendship Network Distribution

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



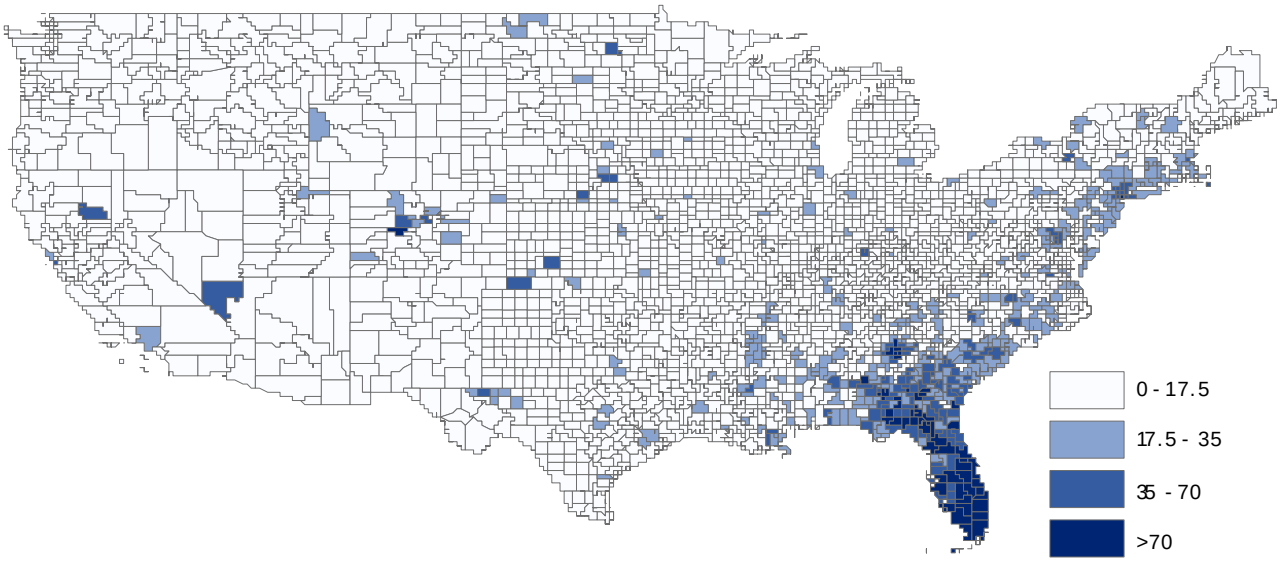
(B) Relative Probability of Friendship Link to Vernon Parish, LA ($RelativeProbFriendship_{i,j}$)



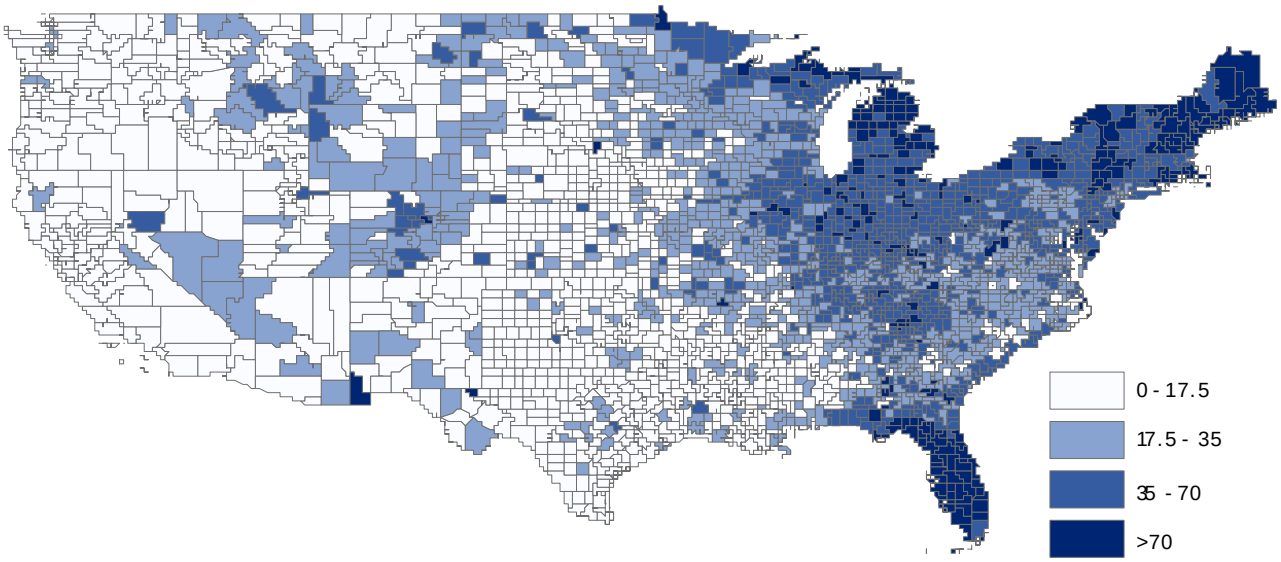
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Rapides Parish, LA in Panel G, and Vernon Parish, LA in Panel H. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Rapides or Vernon) and county j.

Figure A18: Florida Retirement Communities and Friendship Network Distribution

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



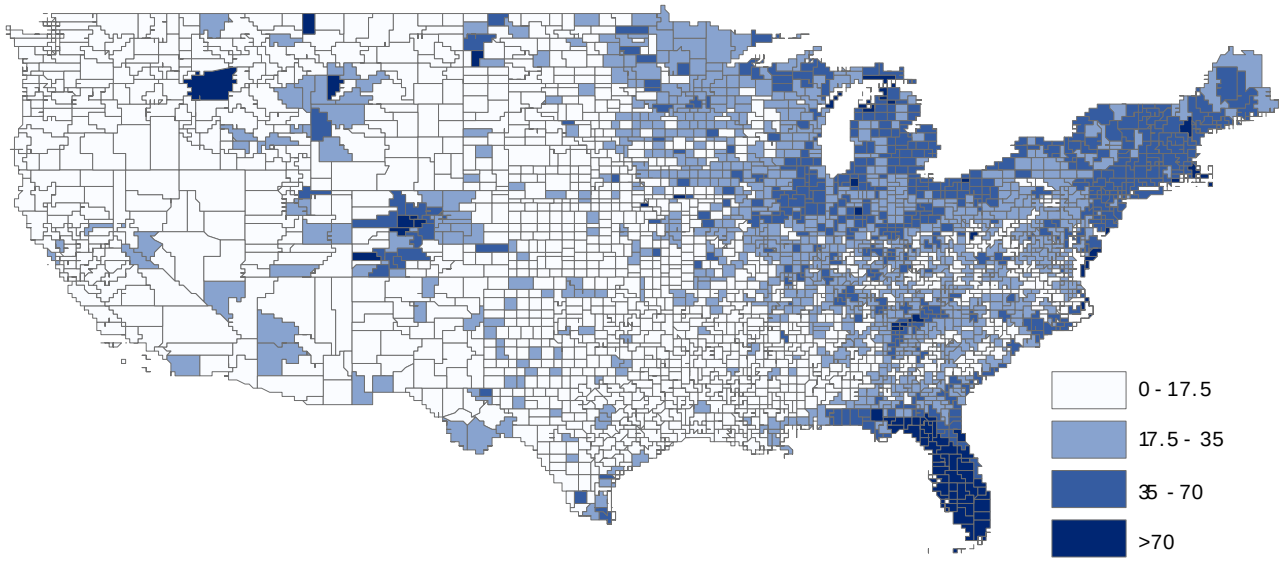
(B) Relative Probability of Friendship Link to Charlotte County, FL ($RelativeProbFriendship_{i,j}$)



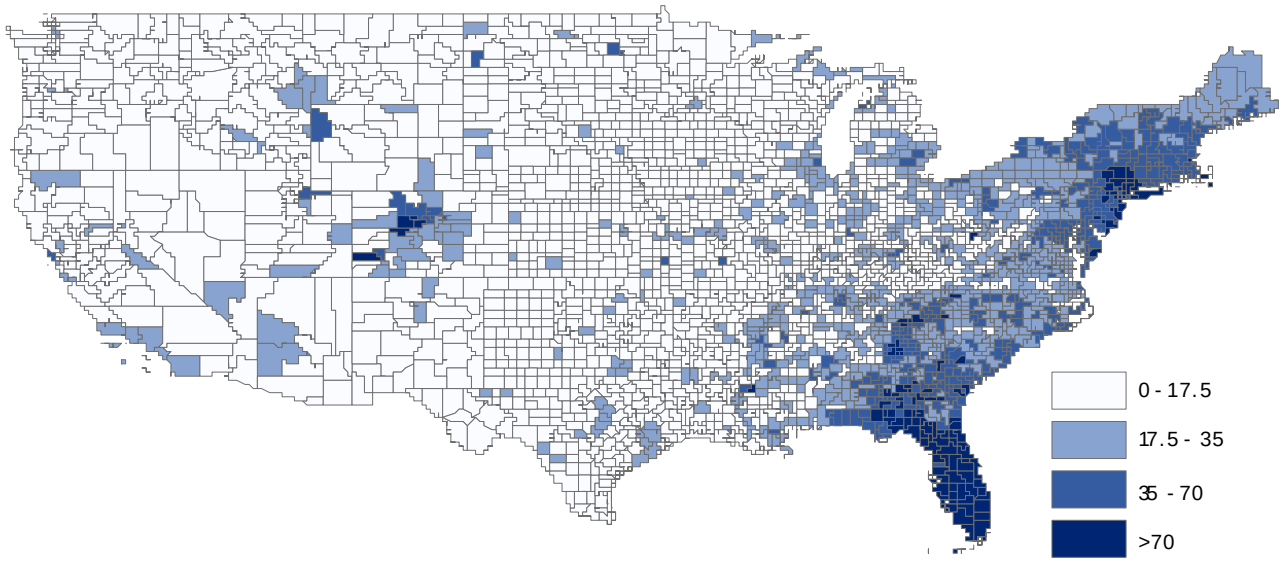
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Miami-Dade County, FL in Panel A, and Charlotte County, FL in Panel B. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Miami-Dade or Charlotte) and county j.

Figure A18: Florida Retirement Communities and Friendship Network Distribution

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



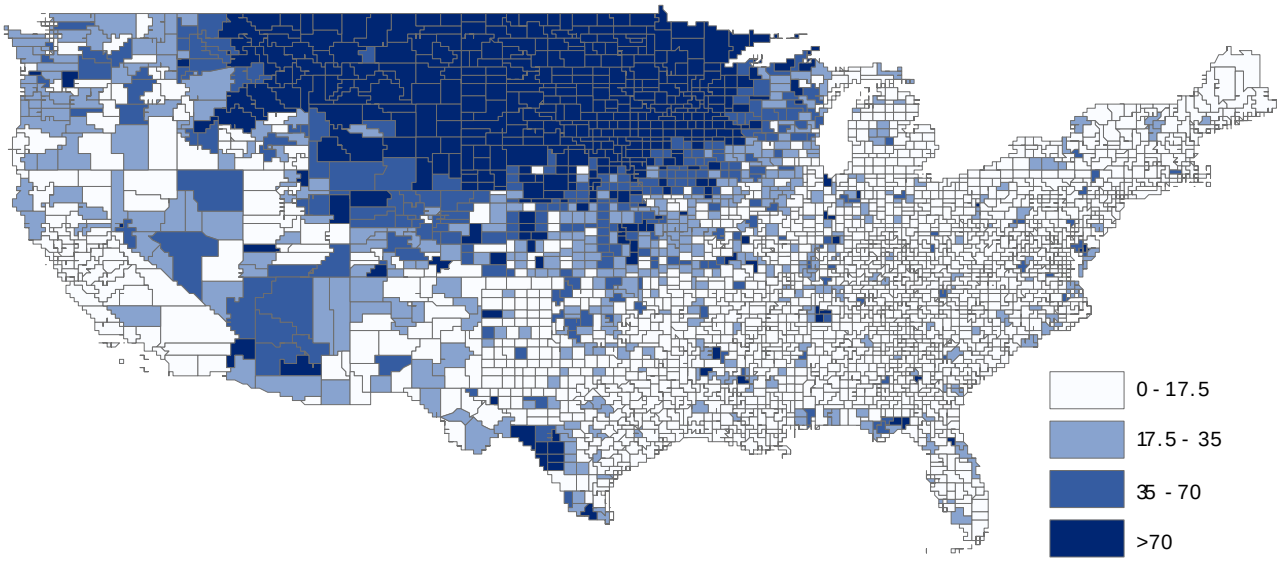
(D) Relative Probability of Friendship Link to Palm Beach County, FL ($RelativeProbFriendship_{i,j}$)



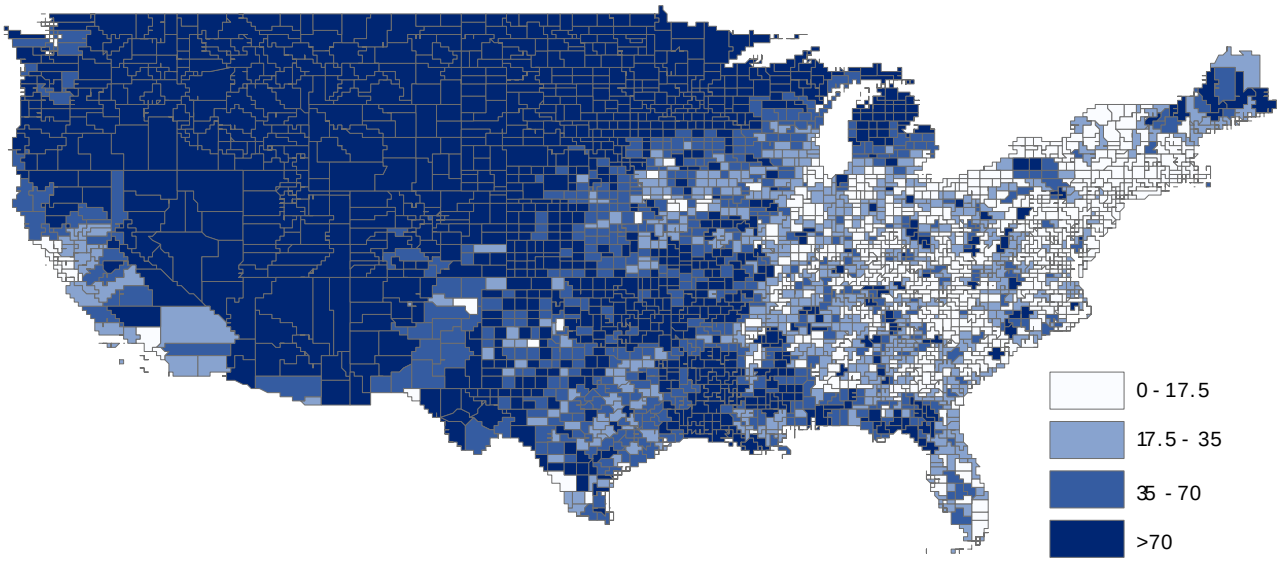
Note: Figure shows the relative probability that a Facebook user in each county j has a friendship link to Collier County, FL in Panel C, and Palm Beach County, FL in Panel D. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Collier or Palm Beach) and county j.

Figure A19: North Dakota Shale Oil Boom and Friendship Network Distribution

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



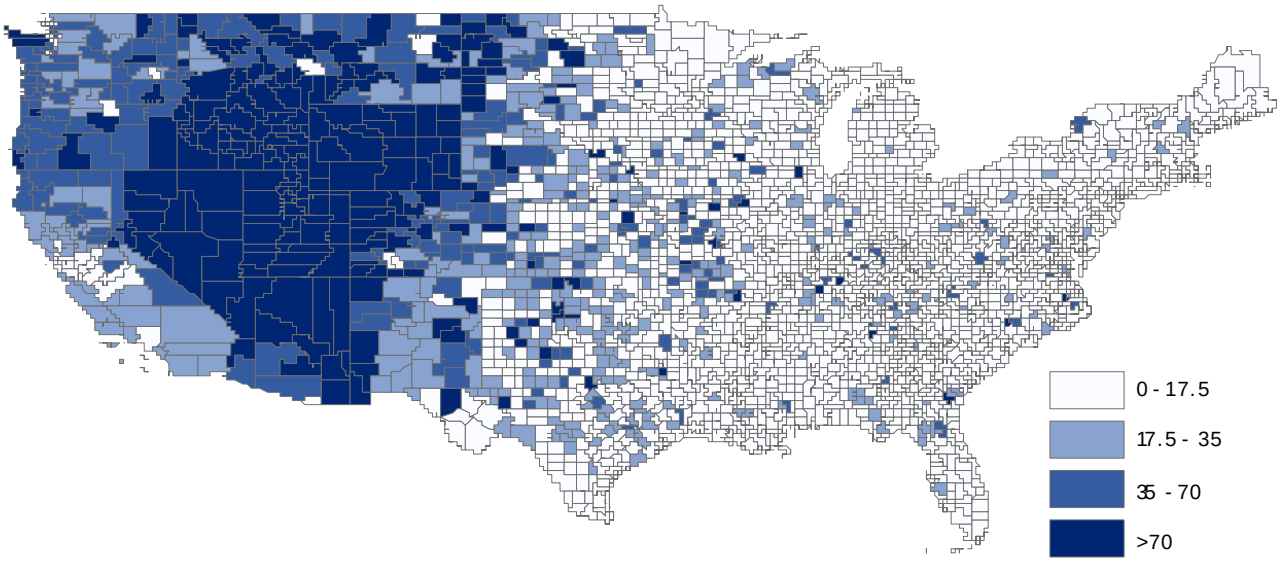
(B) Relative Probability of Friendship Link to McKenzie County, ND ($RelativeProbFriendship_{i,j}$)



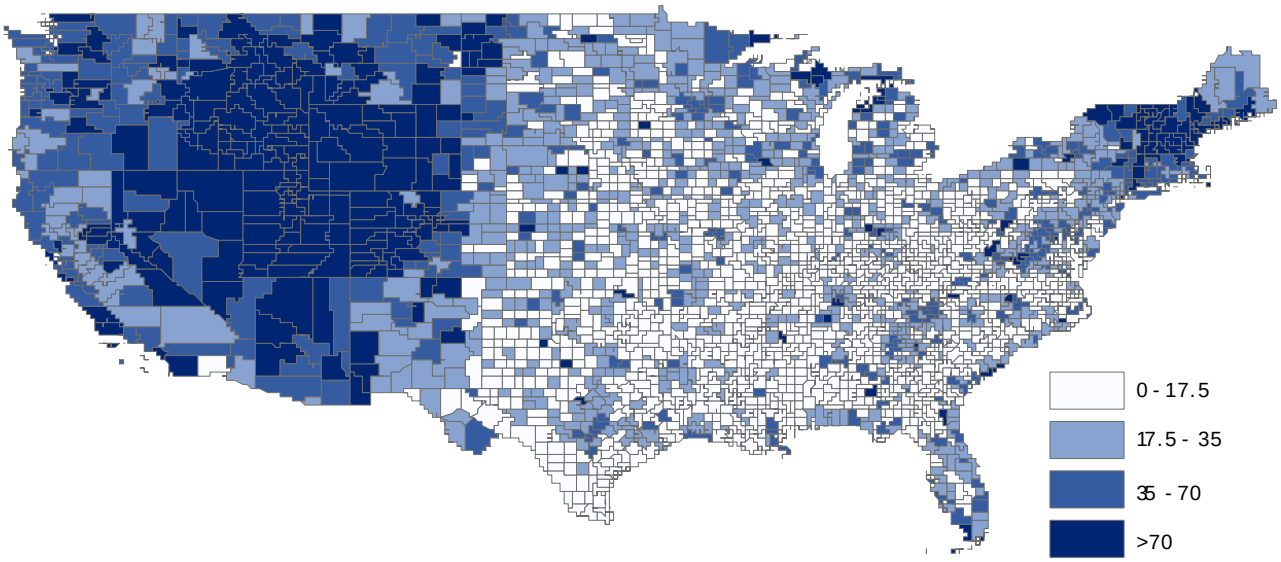
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Richlands County, ND in Panel A, and McKenzie County, ND in Panel B. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Richlands or McKenzie) and county j.

Figure A20: Linkages Between Geographically Distant Winter Sports Areas

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



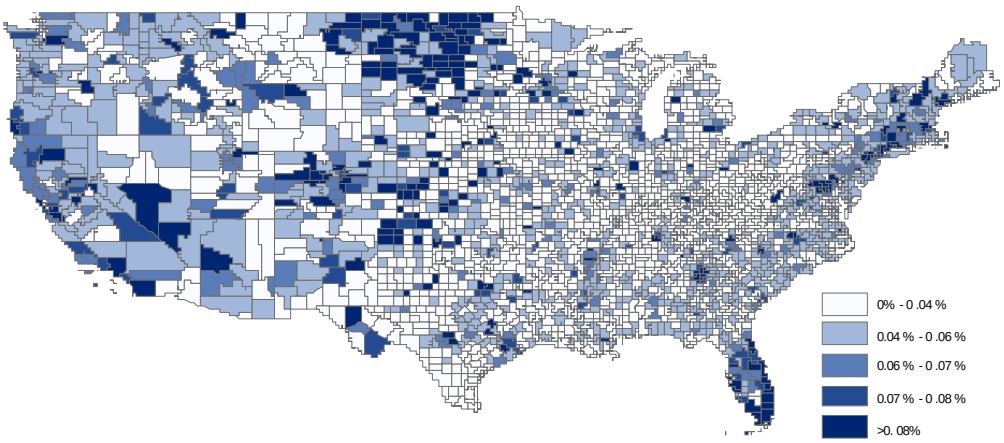
(B) Relative Probability of Friendship Link to Summit County, UT ($RelativeProbFriendship_{i,j}$)



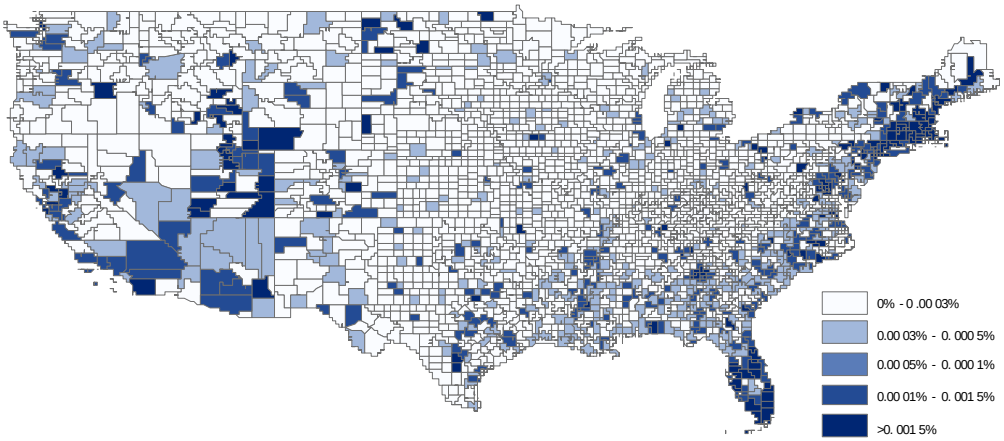
Note: Figure shows the scaled relative probability that a Facebook user in each county j has a friendship link to Sanpete County, UT in Panel A, and Summit County, UT in Panel B. It is constructed as in equation 2. Darker colors correspond to counties in which there is a higher probability of a friendship link between a person in home county i (Sanpete or Summit) and county j.

Figure A21: International Social Connectedness

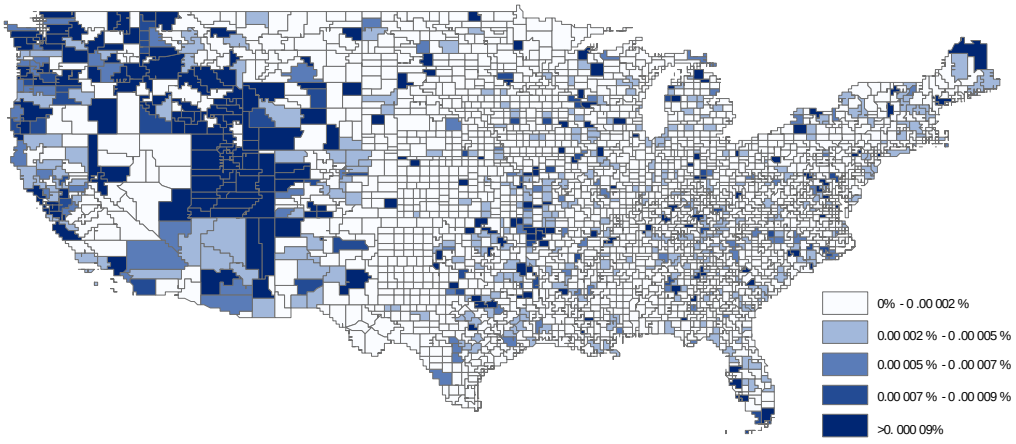
(A) South Africa



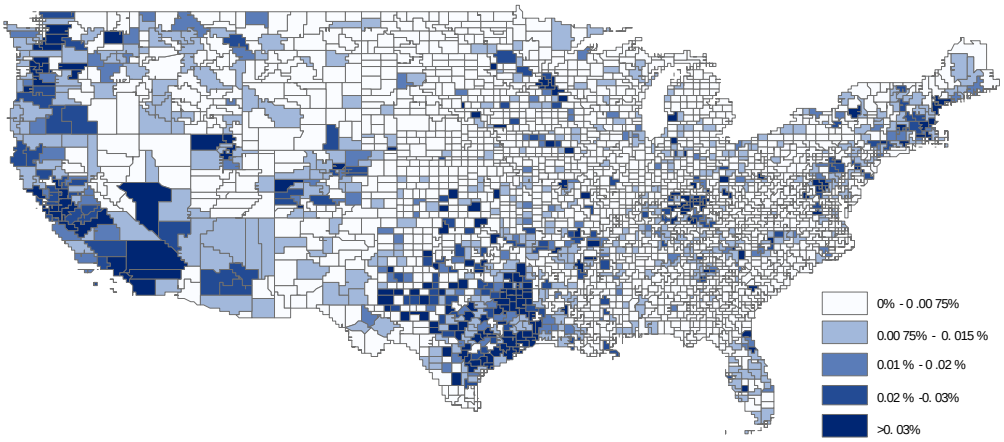
(B) Cape Verde



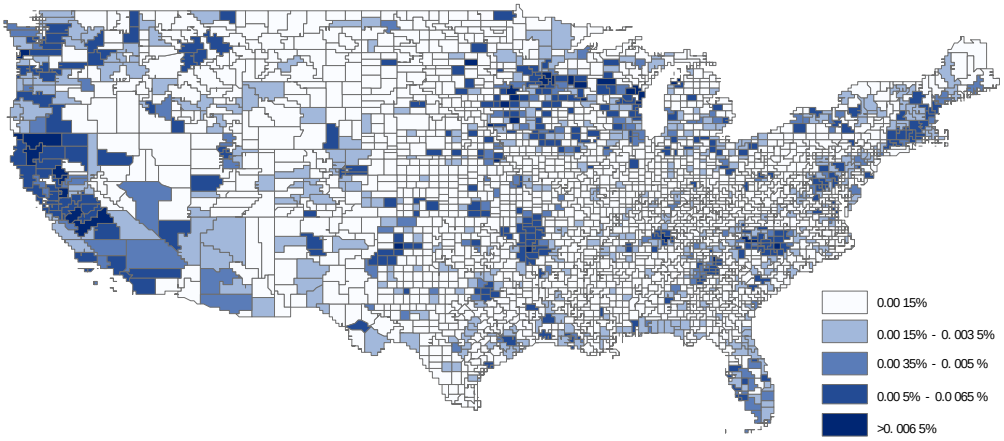
(C) Kiribati



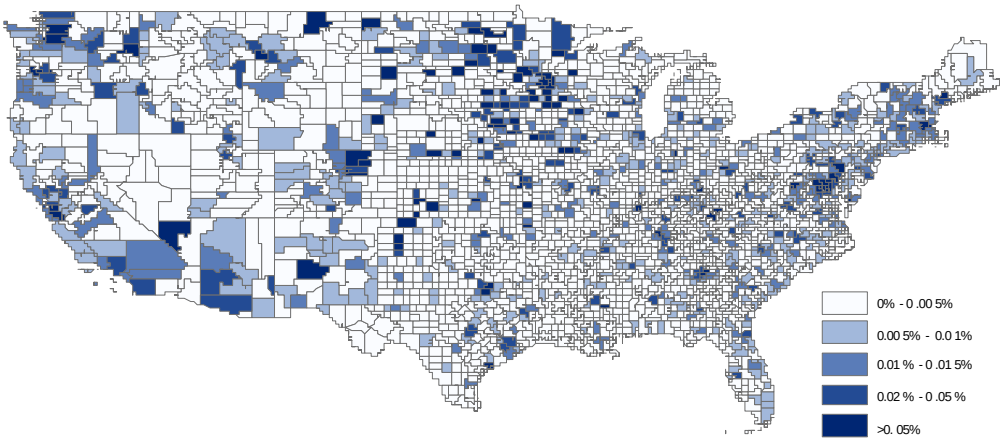
(D) Cambodia



(E) Laos



(F) Ethiopia



Note: Figure shows a heatmap of the share of friendship links that are to Facebook users located in South Africa (Panel A), Cape Verde (Panel B), Kiribati (Panel C), Cambodia (Panel D), Laos (Panel E), and Ethiopia (Panel F).