

## **Social Connectedness: Measurement, Determinants, and Effects**

Michael Bailey, Rachel Cao, Theresa Kuchler,  
Johannes Stroebel, and Arlene Wong

**S**ocial networks can shape many aspects of social and economic activity: migration and trade, job-seeking, innovation, consumer preferences and sentiment, public health, social mobility, and more. In turn, social networks themselves are associated with geographic proximity, historical ties, political boundaries, and other factors. Traditionally, the unavailability of large-scale and representative data on social connectedness between individuals or geographic regions has posed a challenge for empirical research on social networks. More recently, a body of such research has begun to emerge using data on social connectedness from online social networking services such as Facebook, LinkedIn, and Twitter. To date, most of these research projects have been built on anonymized administrative microdata from Facebook, typically by working with coauthor teams that include Facebook employees. However, there is an inherent limit to the number of researchers that will be able to work with social network data through such collaborations.

In this paper, we therefore introduce a new measure of social connectedness at the US county level. Our Social Connectedness Index is based on friendship links on Facebook, the global online social networking service. Specifically, the Social Connectedness Index corresponds to the relative frequency of Facebook friendship

■ *Michael Bailey is a Senior Economic Research Scientist, Facebook, Inc., Menlo Park, California. Rachel Cao is a PhD candidate in economics, Harvard University, Cambridge, Massachusetts. Theresa Kuchler is an Assistant Professor of Finance and Johannes Stroebel is a Professor of Finance, both at the Stern School of Business, New York University, New York City, New York. Arlene Wong is an Assistant Professor of Economics at Princeton University, Princeton, New Jersey. Stroebel is the corresponding author at johannes.stroebel@nyu.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.32.3.259>

doi=10.1257/jep.32.3.259

links between every county-pair in the United States, and between every US county and every foreign country. Given Facebook's scale, with 2.1 billion active users globally and 239 million active users in the United States and Canada (Facebook 2017), 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. Moreover, the Social Connectedness Index data can be made accessible to members of the broader research community. Interested researchers are invited to email [sci\\_data@fb.com](mailto:sci_data@fb.com) to learn about the current process for working with the Social Connectedness Index data.

We begin this article by describing the construction of the Social Connectedness Index (SCI). The bulk of the paper then explores various patterns related to social connectedness. We first use the SCI data to analyze patterns of social connectedness between US 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. We also look at how social connectedness is shaped by political boundaries such as state lines, exposure to large within-US population movements, and other historical and contemporaneous factors.

We then explore heterogeneity across counties in the geographic concentration of their populations' social networks. For the average county, 62.8 percent of all friendship links are to individuals living within 100 miles, but this number ranges from 46.0 percent at the 5th percentile to 76.9 percent at the 95th percentile of the across-county distribution. We find that the populations of counties with a larger fraction of friends living more than 100 miles away are on average better off along a number of socioeconomic dimensions. For example, counties with more geographically dispersed social networks have higher incomes, higher education levels, and higher social mobility.

We then turn to the question of how the intensity of social connectedness between regions correlates with bilateral economic and social activity. We first document a strong correlation between social connectedness and trading activity, consistent with recent research that argues that social networks help overcome informational and cultural frictions that can inhibit trade. Social connectedness is also positively correlated with the spread of innovation and within-US migration. When we look at friendship links between US regions and foreign countries, we find further strong correlations with both past migration patterns and present-day trade flows.

Throughout this essay, our focus is on documenting and describing salient patterns of social connectedness across a variety of settings. We do not seek to provide causal analyses, nor do we want to imply causal relationships behind the correlations we document. Nevertheless, we do believe that our findings can guide future research on the causal effects of social networks. More generally, the patterns discussed here highlight significant opportunities for using data from online social networking services such as Facebook to help alleviate the measurement challenges faced by researchers across the social sciences trying to better understand the role of social connectedness.

## Measuring Social Connectedness

The Social Connectedness Index is constructed using aggregated and anonymized information from the universe of friendship links between all Facebook users as of April 2016. Duggan, Ellison, Lampe, Lenhart, and Madden (2015) report that as of September 2014, more than 58 percent of the US adult population and 71 percent of the US online population used Facebook. The same source reports that, among online US adults, Facebook usage rates are relatively constant across income groups, education groups, and racial groups. Usage rates among online US adults are declining in age, from 87 percent of 18-to-29 year-olds to 56 percent of above-65 year-olds.

In the United States, 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, Goulet, Rainie, and Purcell 2011). Establishing a friendship link on Facebook requires the consent of both individuals, and the total number of friends for a person is limited to 5,000. As a result, Facebook data have a unique ability to provide a large-scale representation of US friendship networks.

To measure the social connectedness between geographies, we map Facebook users to their respective county and country locations, and thus obtain the total number of friendship links between these geographies. Locations are assigned to users based on the users' information and activity on Facebook, including the stated city on their Facebook profile, and device and connection information. We only consider friendship links among Facebook users who have interacted with Facebook over the 30 days prior to the April 2016 snapshot.<sup>1</sup> We treat each friendship link identically.

We then construct the Social Connectedness Index between all pairs of 3,136 US counties, and between every US county and every foreign country, as the normalized total number of friendship links for each geographic pair. In particular, the Social Connectedness Index is constructed to have a maximum value of 1,000,000, and relative differences in the index correspond to relative differences in the total number of friendship links. The highest Social Connectedness Index value of 1,000,000 is assigned to Los Angeles County–Los Angeles County connections (Los Angeles County is where people have the most friends with other people in their county).

## The Determinants of Social Connectedness

The Social Connectedness Index can be used to analyze the correlates of the intensity of social connectedness between US counties. We first analyze the role

<sup>1</sup>Facebook formally defines such “monthly active users” in its 10Q statements as follows: “We define a monthly active user as a registered Facebook user who logged in and visited Facebook through our website or a mobile device, or used our Messenger application (and is also a registered Facebook user), in the last 30 days as of the date of measurement.”

of geographic distance in shaping social connectedness in the United States. The effects of geographic proximity on friendship formation and social interactions have been studied in a number of papers, including Zipf (1949), Verbrugge (1983), and Marmaros and Sacerdote (2006).

As a motivating example, compare San Francisco County and Kern County in California. These two counties have roughly the same population of slightly under one million, but Kern County is 175 times larger in area. Moreover, San Francisco County, which is home to the city of San Francisco, is surrounded by the urbanized Bay Area economy including Oakland and San Jose. Kern County includes the Bakersfield metro area, but it is not surrounded by an urban area.

We construct a measure that we call the “relative probability of friendship” by taking the Social Connectedness Index between counties  $i$  and  $j$  and dividing it by the product of the number of Facebook users in the two counties. This allows us to take into account the fact that we will see more friendship links between counties with more Facebook users.<sup>2</sup> If this measure is twice as large, this means that a given Facebook user in county  $i$  is about twice as likely to be connected with a given Facebook user in county  $j$ . The heat maps in Figure 1 show the relative probability that a given Facebook user in San Francisco County (Figure 1A) or Kern County (Figure 1B) is connected to a given Facebook user in another county.

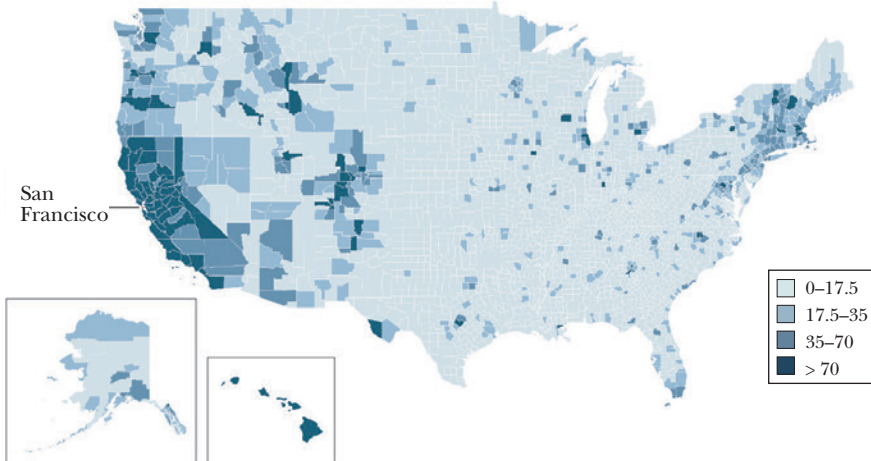
For both San Francisco County and Kern County, a significant proportion of friendship links (dark shading indicates more links) are to geographically close counties across the West Coast. However, there are also noticeable differences in the social connectedness of the two counties. The population of San Francisco County has significant social connections to counties located in the northeastern United States, while 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, because Kern County was a major destination for migrants fleeing the Dust Bowl in the 1930s. Kern County also has substantial friendship links to 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.

Overall, 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 percent of friends living within 50 miles, relative to 27 percent for San Francisco County. In comparison with the summary statistics for the whole United States, displayed in Table 1, the geographic concentration of the friendship network of Kern County is similar to the US average while San Francisco County’s friendship network is extremely geographically dispersed. For the average (population-weighted) US county, 55.4 percent of friends live within 50 miles, with a

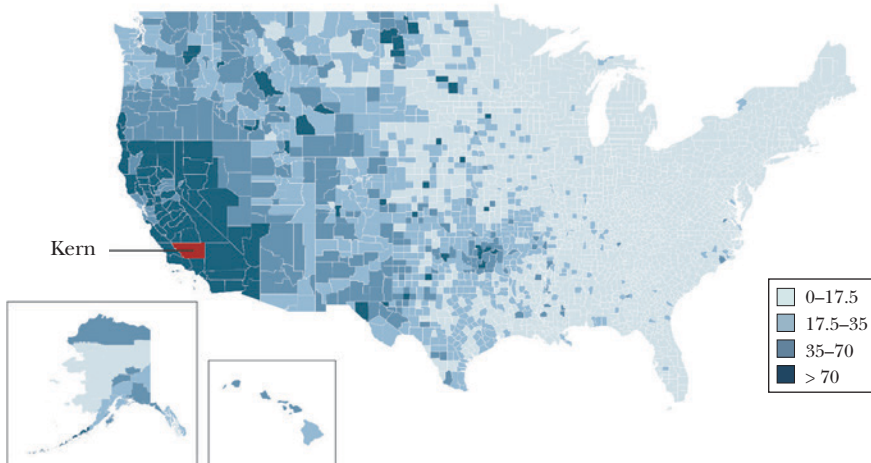
<sup>2</sup>While the number of Facebook users per county is not part of the public data release, very similar patterns for “relative probability of friendship” would be obtained if we instead divided the Social Connectedness Index by the product of county-level populations.

Figure 1  
**County-Level Friendship Maps**

A: Relative Probability of Friendship Link to San Francisco County, CA



B: Relative Probability of Friendship Link to Kern County, CA



*Note:* The heat maps show the relative probability that a Facebook user in each county  $j$  has a friendship link to San Francisco County, CA (Panel A) and Kern County, CA (Panel B). 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  $j$ . The “relative probability of friendship” is constructed by taking the Social Connectedness Index between counties  $i$  and  $j$  and dividing it by the product of the number of Facebook users in the two counties.

10–90 percentile range of 42.5 to 67.4 percent; and over 70 percent of friends live within 200 miles, with a 10–90 percentile range of 57.1 to 81.2 percent. This despite the fact that, for the average county, only 1.3 percent and 6.6 percent of the US population live within 50 miles and 200 miles, respectively.

Table 1

**Distance and Friendship Links: Across-County Summary Statistics for the United States**

	<i>Share of friends living within:</i>			<i>Share of US population living within:</i>		
	<i>50 Miles</i>	<i>100 Miles</i>	<i>200 Miles</i>	<i>50 Miles</i>	<i>100 Miles</i>	<i>200 Miles</i>
Mean	55.4%	62.8%	70.3%	1.3%	2.8%	6.6%
P5	38.1%	46.0%	54.2%	0.1%	0.3%	1.0%
P10	42.5%	49.6%	57.1%	0.1%	0.6%	2.1%
Median	55.4%	63.9%	71.6%	0.7%	2.1%	5.8%
P90	67.4%	74.8%	81.2%	3.2%	6.2%	15.0%
P95	70.3%	76.9%	83.2%	5.4%	9.2%	15.6%

*Note:* Table shows across-county summary statistics for the share of friends of a county's population living within a certain distance of that county as well as the share of the US population living within those distances. P5, P10, P90, and P95 are the 5th, 10th, 90th, and 95th percentiles, respectively. Counties are weighted by their populations.

The regressions in Table 2 offer a more systematic account of the relationship between geographic distance and social connectedness across county-pairs. The unit of observation is a county-pair. The dependent variable is the log of the Social Connectedness Index between the two counties. The log of the geographic distance between the counties is the explanatory variable in column 1. We include fixed effects for both counties, which controls for population levels and any other characteristics that vary at the county level. In this specification, geographic distance is able to explain a significant amount of the cross-county-pair variation in social connectedness. The estimated elasticity of social connectedness to geographic distance suggests that a 10 percent increase in the distance between two counties is associated with a 14.8 percent decline in the number of friendship links between those counties. Similar to gravity equations estimated in the trade literature, this estimates the equilibrium relationship between geographic distance and social connectedness, not necessarily the causal effect of one on the other.

In column 2, we include an additional control indicating whether 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. This finding is not the result of non-log linearities in the distance relationship, and it can be found for both border counties and nonborder counties (as we discuss further in the Appendix). Why social connectedness varies so strongly at state borders, and the extent to which this is driven by institutional, social, or economic factors, is an interesting avenue for future research. Possible explanations include the importance of common state-level identities or the role of state universities as meeting places for residents from the same state.

In columns 3 and 4, 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 estimated elasticity between geographic distance and friendship links is  $-1.99$ . 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 suggest 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. In turn, this pattern highlights that in the theoretical modeling of friendship links, the appropriate elasticity depends on the geographic distances studied. This finding may help to explain why previous estimates of the elasticity of friendship probability with respect to geographic distance vary so significantly across settings, including an estimate of  $-2$  in a study of cell-phone communication networks in the United Kingdom (Lambiotte et al. 2008); an estimate of  $-1$  among bloggers (Liben-Nowell, Novak, Kumar, Raghavan, and Tomkins 2005); and an estimate of  $-0.5$  in location-based online social networks such as Brightkite, Foursquare, and Gowalla (Scellato, Noulas, Lambiotte, and Mascolo 2011).

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 (for a comprehensive review of the literature, see McPherson, Smith-Lovin, and Cook 2001). Thus, in column 5 of Table 2 we add a number of variables measuring the similarity of counties on measures such as per capita income, education levels, and religiosity. We find that county pairs that are more similar on these dimensions have more friendship links. However, while the magnitude of the effect of these socioeconomic differences on social connectedness is potentially meaningful, adding them barely affects the coefficients on other explanatory variables or the  $R^2$  relative to the specification in column 2.

Table 2 highlights that social connectedness drops off strongly at state borders. A related question is how closely the existing state borders resemble the borders that would form if we grouped together US counties to create communities with the aim of maximizing within-community social connectedness. There are a number of possible algorithms to facilitate such a grouping of counties. Here, we use a method called hierarchical agglomerative linkage clustering (which we describe further in the online Appendix).

Figure 2 shows the result when we use this algorithm to group the United States into 20 distinct communities. All resulting communities are spatially contiguous, which 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. Another group of states is Florida, Georgia, and Alabama. However, some states are split into separate communities. The Texas panhandle is grouped with Oklahoma and Kansas, and Colorado’s Western Slope forms its own community.

Table 2

**Determinants of Social Connectedness across County Pairs**

	Dependent Variable: Log(SCI)				
	(1)	(2)	(3)	(4)	(5)
log(Distance in Miles)	-1.483*** (0.065)	-1.287*** (0.061)	-1.160*** (0.059)	-1.988*** (0.043)	-1.214*** (0.055)
Same State		1.496*** (0.087)	1.271*** (0.083)	1.216*** (0.044)	1.496*** (0.085)
$\Delta$ Income (\$1,000)					-0.006*** (0.001)
$\Delta$ Share Population White (%)					-0.012*** (0.001)
$\Delta$ Share Population No High School (%)					-0.012*** (0.002)
$\Delta$ 2008 Obama Vote Share (%)					-0.006*** (0.001)
$\Delta$ Share Population Religious (%)					-0.002*** (0.001)
County Fixed Effects	Y	Y	Y	Y	Y
Sample			>200 miles	<200 miles	
Number of observations	2,961,968	2,961,968	2,775,244	186,669	2,961,968
$R^2$	0.907	0.916	0.916	0.941	0.922

*Note:* Table shows results from a regression of the log of the Social Connectedness Index on a number of explanatory variables. The log of the geographic distance between the counties is the explanatory variable in column 1. In column 2, we include an additional control indicating whether both counties are within the same state. In columns 3 and 4, we restrict the sample to county-pairs that are more and less than 200 miles apart, respectively. The unit of observation is a county-pair. Standard errors are given in parentheses. The online Appendix (<http://e-jep.org>) provides more details on the data sources and exact specifications.

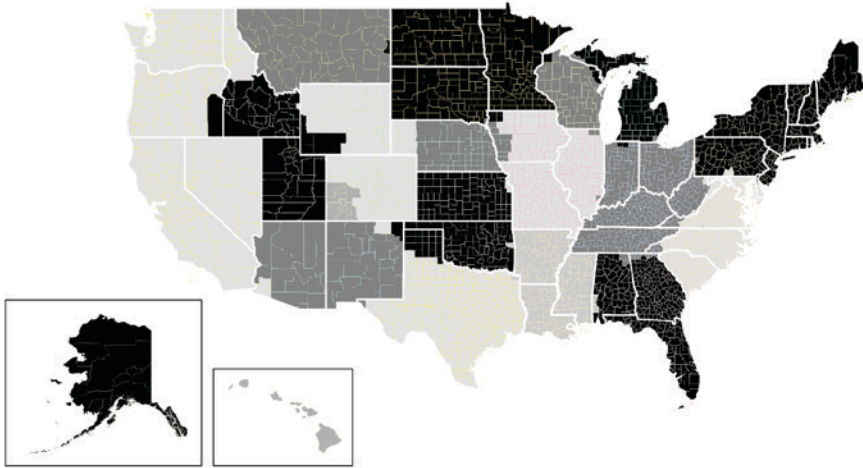
\*, \*\*, and \*\*\* indicate significance levels of  $p < 0.1$ ,  $p < 0.05$ , and  $p < 0.01$ , respectively.

These findings suggest that it might be interesting to study the economics and politics of US “regions” as defined by joint social connectedness, rather than alternative groupings such as Census regions or divisions.

We have explored a number of additional correlates of friendship links across counties. For example, we document that the strength of social connections can 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. Counties with Native American reservations are strongly connected to one another. Similarly, areas with ski resorts in the Rocky Mountains and New England have high social connectedness. Counties in Florida with significant retiree populations are strongly connected to the Rust Belt and the Northeast. In addition, large cities in the Midwestern



Figure 2

**Connected Communities within the United States—20 Units**

*Note:* Figure shows US counties grouped together when we use hierarchical agglomerative linkage clustering to create 20 distinct groups of counties.

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. For more details on these patterns, see the online Appendix (<http://e-jep.org>). In general, many of these patterns of friendship connections are unsurprising, but it is new that such patterns can now be measured and documented in systematic national data.

### **Concentration of Social Networks and County Characteristics**

The geographic concentrations of the friendship networks of different counties reveal a great deal of heterogeneity: for example, the earlier Table 1 shows that the 5th–95th percentile range across population-weighted counties in the share of friends living within 100 miles is 46.0 percent to 76.9 percent. Existing theoretical work suggests that the diversity of social networks is an important determinant of economic development; conversely, tightly clustered social ties can limit access to a broad range of social and economic opportunities (for example, Granovetter 1973). However, empirical studies of the relationship between the structure of social networks and economic outcomes of communities are rare. One exception is Eagle, Macy, and Claxton (2010), who use UK cellphone 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 across US counties is highly correlated with social and economic

outcomes at the county level, such as average income, educational attainment, and social mobility.

If we define the concentration of a friendship network as the share of friends who live within 100 miles, then friendship networks in the South, the Midwest, and Appalachia are the most geographically concentrated. Counties in the Rocky Mountains have the smallest share of friends living within 100 miles, in large part because these areas are often less-densely populated. Among the western United States, Utah and inland California have the most geographically concentrated friendship networks. The online Appendix shows heat maps of this and other measures of the geographic concentration of friendship networks.

What are the effects of differentially structured social networks on county-level outcomes? As a first step toward answering this question, we correlate our measure of the concentration of friendship links with county-level characteristics. Figure 3 presents county-level binned scatterplots using the share of friends living within 100 miles and a number of socioeconomic outcomes. The overall message is that counties where people have more concentrated social networks tend to have worse socioeconomic outcomes along a number of dimensions: on average, they have lower income, lower education, higher teenage birth rate, lower life expectancy, less social capital, and less social mobility.

These correlations cannot be interpreted as causal (although the online Appendix discusses a number of causal mechanisms proposed by the literature that are consistent with our findings). Our goal here, as in the rest of the paper, is to document patterns that can guide future research investigating the causal effects of social network structure on socioeconomic outcomes, and to describe the Social Connectedness Index data that can help with such analyses. 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.

## **Social Connectedness and Cross-County Activity**

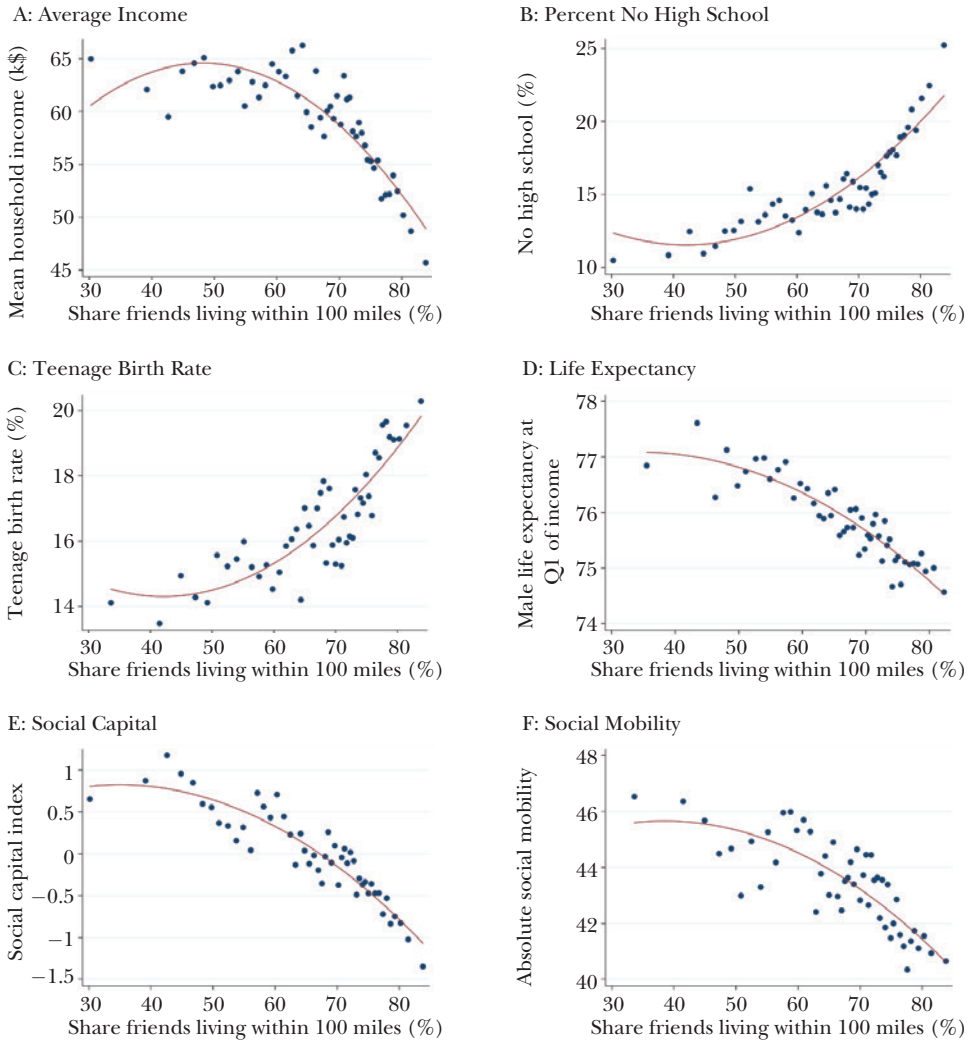
Social connectedness between two regions may be related to other economic and social interactions between these regions. Indeed, we next document correlations between the number of friendship links and trade flows, patent citations, and migration patterns. As before, we illustrate some salient patterns in the data rather than providing full-fledged causal analyses. For each of the patterns documented below, the online Appendix (<http://e-jep.org>) provides more details on the variables, data construction, specifications, and additional exploration.

### **Social Connectedness and Within-US Trade Flows**

A well-established empirical result in the trade literature is that bilateral trade between two regions decreases with geographic distance, although the explanations

Figure 3

Network Concentration and County-Level Characteristics



Notes: Panels show binned scatterplots with counties as the unit of observation. To generate each binned scatterplot, we group the x-axis variable into 50 equal-sized bins. We then compute the mean of the x-axis and y-axis variables within each bin and create a scatterplot of these 50 data points. The horizontal axes measure the share of friends of the county population that live within 100 miles. On the vertical axes are a number of county-level measures of socioeconomic outcomes: the mean county income in Panel A; the share of the population with no high school degree in Panel B; the teenage birth rate as provided by Chetty, Hendren, Kline, and Saez (2014) in Panel C; the life expectancy of males in the first quarter of the national income distribution from Chetty et al. (2016) in Panel D; the measure of social capital in 2009 as defined by Rupasingha, Goetz, and Freshwater (2006) in Panel E; and the absolute measure of social mobility from Chetty et al. (2014) in Panel F. The red line shows the fit of a quadratic regression. The online Appendix (<http://e-jep.org>) provides more details.

for this finding are still being debated (for a review, see Anderson and van Wincoop 2004). Many studies have highlighted that the distance effect is too large to be fully explained by trade costs alone, and that geographic distance might serve as a proxy for other trade frictions such as cultural differences, lack of familiarity, or information asymmetries. Social connections may alleviate the trade costs associated with these factors, and some empirical work has examined the causal effect of stronger social networks on trade (Rauch 1999; Combes, Lafourcade, and Mayer 2005; 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 thus had to rely on indirect proxies, such as the ethnic composition of regions or past migration patterns.

The Social Connectedness Index data allow us to examine directly the empirical relationship between trade flows and social connectedness at the state level. Panel A of Table 3 shows some results. For the dependent variable, we measure interstate trading volumes using data from the Commodity Flow Survey. We focus on data from 2012, the latest year with comprehensively available data. Specifically, the dependent variable captures the log of the value of trade in 2012 between origin state  $i$  and destination state  $j$ .

For our main explanatory variables, we use the log of geographic distance between states  $i$  and  $j$ , as well as the log of the Social Connectedness Index between states  $i$  and  $j$  (constructed from a weighted average of county-level SCI measures). We also include fixed effects for each state, dummy variables for own-state flows, and dummy variables if the states are adjacent to each other.

We observe two main patterns. First, social connectedness is strongly correlated with state–state trade flows, even after controlling for geographic distance. The magnitude of the elasticity of trade with social connectedness is large and statistically significant.<sup>3</sup> In fact, when comparing them 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 elasticity of trade halves in magnitude after controlling for social connectedness. In column 4, we further control for differences across the states in GDP per capita, unemployment rates, sectoral composition, union share, and population density. The addition of these further controls has essentially no effect on the estimated elasticity between social connectedness and trade.

The observed reduction in the distance elasticities of trade, after controlling for social connectedness, is consistent with theories described above which suggest that geographic distance might be proxying for other factors affecting trade between

<sup>3</sup>In the online Appendix, we explore these patterns across industries. We find that the magnitude of the elasticity of trade flows with respect to friendship links rises with the share of high-skilled workers in the sector and is not affected by the share of labor compensation in total costs.

Table 3

**Social Connectedness and Across-Region Economic Interactions**

	(1)	(2)	(3)	(4)
<i>Panel A: Dependent Variable: log(State-Level Trade Flows)</i>				
log(Distance)	-1.057*** (0.071)		-0.531*** (0.084)	-0.533*** (0.085)
log(SCI)		0.999*** (0.051)	0.643*** (0.071)	0.637*** (0.060)
State Fixed Effects	Y	Y	Y	Y
Other State Differences	N	N	N	Y
Observations	2,219	2,220	2,219	2,219
R <sup>2</sup>	0.912	0.918	0.926	0.930
<i>Panel B: Dependent Variable: Indicator for Patent Citation</i>				
log(Distance)	-0.048*** (0.002)		-0.011** (0.005)	-0.021** (0.009)
log(SCI)		0.063*** (0.003)	0.049*** (0.006)	0.066*** (0.012)
Technological Category + County Fixed Effects	Y	Y	Y	Y
Cited + Issued Patent Fixed Effects, Other County Differences	N	N	N	Y
Observations	2,171,754	2,171,754	2,171,754	2,168,285
R <sup>2</sup>	0.056	0.059	0.059	0.101
<i>Panel C: Dependent Variable: log(County-Level Migration)</i>				
log(Distance)	-0.973*** (0.048)		0.023 (0.021)	0.031 (0.021)
log(SCI)		1.134*** (0.019)	1.148*** (0.024)	1.159*** (0.024)
County Fixed Effects	Y	Y	Y	Y
Other County Differences	N	N	N	Y
Observations	25,305	25,305	25,305	25,287
R <sup>2</sup>	0.610	0.893	0.893	0.893

*Note:* Table shows the relationship between bilateral economic activity across geographic units and the geographic distance and social connectedness between these units. “SCI” stands for Social Connectedness Index. In Panel A, the unit of observation is a state-pair, and the dependent variable is the log of the value of 2012 trade flows between the states. All specifications include state fixed effects, dummies for own state, and dummies for neighboring states; column 4 also controls for differences across states on important socioeconomic indicators. In Panel B, the unit of observation is a patent-pair. The dependent variable is an indicator of whether patent  $i$  cites patent  $j$ . All specifications control for the county and technology category fixed effects, and column 4 also controls for patent fixed effects and other differences across the counties of the patents on important socioeconomic indicators. In Panel C, the unit of observation is a county pair, and the dependent variable is the log of across-county migration between 2013 and 2014. All specifications control for county fixed effects, and column 4 also controls for other differences across counties on important socioeconomic indicators. Standard errors are given in parentheses. The online Appendix (<http://e-jep.org>) provides more details on the data sources and exact specifications.

\*, \*\*, and \*\*\* indicate significance levels of  $p < 0.1$ ,  $p < 0.05$ , and  $p < 0.01$ , respectively.

states. Further investigating the causal role of social connectedness in facilitating trade flows might therefore be a useful avenue for future research.

### **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 might therefore have important effects on economic activity, by facilitating the diffusion of knowledge and ideas through society.<sup>4</sup> However, testing these theories is challenging, because both knowledge spillovers and the degree of social connectedness are hard to measure. To overcome these challenges, a large empirical literature has relied on patent citations as a measure of knowledge spillovers (Jaffe, Trajtenberg, and 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. In turn, this finding is often interpreted as evidence for the importance 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).

The Social Connectedness Index has the potential to provide more direct evidence for the role of social connectedness in facilitating knowledge spillovers. We obtain data containing information on all patents granted by the US Patent and Trademark Office in the years 2002–2014, and 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 (as in Berkes and Gaetani 2017). 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.

We follow the approach in the existing literature to explore the relationship between social connectedness and patent citations (for example, Jaffe, Trajtenberg, and Henderson 1993). This approach matches each “citing patent” with a “non-citing patent” issued at the same time and in the same technological class to serve as a control, as we will explain below. 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, after controlling for the patent's technological class and the geographic distance between the geographies. The literature has argued that this approach can help to separate knowledge spillovers

<sup>4</sup>For examples, see Jovanovic and Rob (1989), Kortum (1997), Benhabib and Spiegel (2005), Alvarez, Buera, and Lucas (2008), Comin and Hobijn (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 thus how quickly the region adopts a new idea (for instance, Glaeser 1999; Black and Henderson 1999; Moretti 2012).

from correlations that might be induced by patterns in the geographic location of technologically related activities across regions that are connected through social networks.

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

Panel B of Table 3 shows results from our analysis. The dependent variable in the regressions equals one if an issued patent *i* cites patent *j*, and zero otherwise. The first two rows show the coefficients on the log of geographic distance and the log of the Social Connectedness Index between the counties of the issued and cited patents. We include fixed effects for the technology classes and for the counties of patents *i* and *j*.

Comparing columns 1 and 2, social connectedness explains marginally more of the variation in the probability of a patent citation than geographic distance, as the  $R^2$  in column 2 is higher. In terms of economic magnitudes, the probability of a patent citation is 6.3 percentage points higher when the social connectedness between the counties of the issued and cited patents doubles.

In column 3, we jointly estimate the relationship of geographic distance and social connectedness with the probability of a patent citation. 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  $-4.8$  to  $-1.1$  percent.

In column 4, we also control for a host of across-county differences on important socioeconomic indicators: 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 add fixed effects for the cited and the issued patents. If anything, the estimated relationship between social connectedness and patent citation increases somewhat as a result of these further controls.

This finding suggests that the relationship between geographic distance and the probability of patent citation, viewed in isolation, may be partially capturing effects of information flows associated with social connectedness. More generally, our results suggest a significant correlation between social connectedness and knowledge spillovers, innovation, and, ultimately, economic growth. These findings highlight the potential of the Social Connectedness Index data to help uncover possible causal relationships behind these correlations.

### Social Connectedness and Migration

Understanding the factors driving migration patterns is important. For example, within-US migration is one mechanism for equilibrating the US labor market following regional shocks (Blanchard and Katz 1992). 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 (for a review, see Munshi 2016). While a lot of the research has focused on international migration (for example, Moretti 1999), similar forces might be at work in explaining within-US migration.

We find that the Social Connectedness Index has significant explanatory power for migration between regions, beyond what is predicted by geographic distance. Panel C of Table 3 shows some results. The dependent variable captures the log of total migration between counties  $i$  and  $j$  between 2013 and 2014, as measured by the Statistics of Income (SOI) Tax Stats Migration Data provided by the IRS. The key explanatory variables are the log of geographic distance between those counties and the log of the Social Connectedness Index. We also include fixed effects for each county, which allows us to control for the size of its population and other county-level characteristics that might affect the degree of migration.

In column 1 of Table 3, Panel C, we do not include the social connectedness variable. The estimated elasticity of migration to geographic distance is close to  $-1$ . In column 2, we find that the elasticity of migration to social connectedness is slightly larger than 1, with a somewhat higher  $R^2$  than in column 1. In other words, the Social Connectedness Index can explain a larger part of the variation of the migration flows across county-pairs than geographic distance can. In column 3, we control for both the geographic distance and social connectedness between counties. We find that geographic distance adds no additional predictive power compared with column 2. This finding suggests that much of the estimated 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. Column 4 shows that these conclusions are robust to further controlling for other differences across counties on important socioeconomic indicators.

Overall, our results are consistent with stories in which individuals are more likely to move to counties where they already have friends. Such a mechanism could, for example, result in larger cities attracting even more new movers and thereby help explain the very right-tailed city size distribution (Gabaix 1999). Exploring the causal mechanisms behind the observed relationship between social connectedness and migration thus provides an exciting research agenda.

### International Dimension of Social Connectedness of US Counties

US counties vary considerably in the share of social connections to individuals living outside of the United States. For the median county, 4 percent of all



friendship links are to individuals living in foreign countries, but the 10–90 percentile range is 2.3 percent to 8.6 percent, and the 1–99 percentile range is 1.6 percent to 18.7 percent. Some of this variation is straightforward to explain. For example, areas close to the Mexican or the Canadian border have more international connections. Patterns of past immigration matter as well. For example, connections with Norway are particularly strong for those parts of the United States that saw major immigration from Norway in the late 19th and early 20th Centuries, like Wisconsin, Minnesota, and the Dakotas. Similarly, a number of counties in the northeastern United States have strong social connectedness to Italy. For heat maps of social connectedness to these and other countries, see the online Appendix available with this paper at <http://e-jep.org>.

The first three columns in Table 4 illustrate the extent to which past migration from a particular country is correlated with the strength of today's social connectedness of a US county with that country. In these columns, the dependent variable is the Social Connectedness Index between each county and foreign country. For the explanatory variables, geographic distance is measured between each county and the capital city of each foreign country. We use two measures of past migration: the number of residents who claim their primary ancestry as being from a given foreign country and the number of residents in each county who were born in a specific foreign country. The first measure is broader and can, for instance, include US-born individuals with immigrant parents or grandparents. All variables are measured in logs. We also include fixed effects for each county and foreign country.

The first column shows the correlation between geographic distance and international social connectedness: a 1 percent increase in the geographic distance is associated with a 1.2 percent decline in social connectedness. Interestingly, this elasticity is nearly identical to the elasticity of friendship links to geographic distance estimated for the United States for distances greater than 200 miles. The second column shows that a 1 percent increase in the number of residents with ancestry from a given foreign country correlates with an increase in social connections to that country by about one-third of a percent. In column 3, we obtain similar estimates for our second measure of past migration. Across columns 2 and 3, controlling for past migration reduces the estimated effect of geographic distance on social connectedness by between one-third and one-half.

In other regressions presented in the online Appendix, we find that the effect of past migration on today's social connections is stronger for countries from which immigration to the United States occurred more recently, such as Mexico or the Philippines, compared to countries from which immigration peaked earlier, such as Germany or Ireland. For example, the coefficient on a regression like that in column 2 is about 0.13 for counties with immigration waves that peaked pre-1900 or between 1900 and 1930, but more than twice as high for waves that peaked between 1930 and 1990 or for waves that have not yet peaked.

We also sought to estimate the relationship between social connectedness and international trade. Again, we used state-level data on social connectedness

Table 4

**Social Connectedness, Ancestry, and International Trade**

	log(SCI)			log (Exports + 1)	log (Imports + 1)
	(1)	(2)	(3)	(4)	(5)
log(Distance)	-1.159*** (0.258)	-0.690*** (0.162)	-0.493*** (0.174)	-2.092*** (0.391)	-1.627*** (0.378)
log(Ancestry in Foreign Country)		0.341*** (0.022)			
log(Born in Foreign Country)			0.367*** (0.033)		
log(SCI)				0.597*** (0.139)	0.470*** (0.103)
Fixed Effects	Y	Y	Y	Y	Y
Observations	33,146	33,146	16,527	11,015	11,014
R <sup>2</sup>	0.908	0.936	0.943	0.770	0.770
Number of Countries	105	105	52	216	216

*Note:* The table explores the international dimension of social connectedness. In columns 1 to 3, we explore how past migration patterns and geographic distance are correlated with international social connectedness. The unit of observation is a US county–foreign country pair. Each specification also includes fixed effects for the US state and the foreign country, and the dependent variable is the log of the Social Connectedness Index between those units. In columns 4 and 5, we explore how today’s international trading activity is correlated with social connectedness. The unit of observation is a US state–foreign country pair. Standard errors are given in parentheses. The online Appendix (<http://e-jep.org>) provides more details on the data sources and exact specifications.

\*, \*\*, and \*\*\* indicate significance levels of  $p < 0.1$ ,  $p < 0.05$ , and  $p < 0.01$ , respectively.

(by combining the counties of a given state into a population-weighted average), because data on international trade is only available at the state level. Adjusting for geographic distance, (in a specification similar to Table 3, Panel B, column 3), we find that a state with 10 percent higher social connectedness to a given foreign country on average imports 4.7 percent more from this country and exports 6.0 percent more to this country. These findings are highly consistent with our earlier estimates on within-US trade. In the online Appendix for this paper, we provide additional details on these variables and alternative specifications.

## Conclusion

We use data from the global online social networking site Facebook to construct the Social Connectedness Index (SCI). These data provide a new and comprehensive measure of social connectedness between US county pairs, as well as between US counties and foreign countries. The SCI should allow researchers to overcome some 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 to better understand the geographic dimensions of real-world social networks, as well as to document that social connectedness correlates strongly with social and economic activity across regions. While these correlations should not be seen as identifying causal relationships, they provide starting points for investigating a variety of important questions.

A number of recent studies have used data from online social networks, in most cases by including coauthors from Facebook or other social networking services. For example, Gee, Jones, and Burke (2017) and Gee, Jones, Fariss, Burke, and Fowler (2017) use de-identified microdata from Facebook to analyze the role of social networks in the job-finding process. These researchers were able to assess the relative importance of strong and weak ties in helping job seekers find new employment. Social network data from Facebook have also been used to study a range of other topics: the relationship between the size of friendship networks and mortality (Hobbs, Burke, Christakis, and Fowler 2016); the structure of social networks in immigrant communities in the United States (Herdağdelen, State, Adamic, and Mason 2016); the evolution of information cascades (Cheng, Adamic, Kleinberg, and Leskovec 2016); and the effects of social influence and social advertising (Bakshy, Eckles, Yan, and Rosenn 2012). Other researchers have studied the effects of online social networks themselves. For example, Bakshy, Messing, and Adamic (2015) study how online networks influence exposure to perspectives that cut across ideological lines. In our own work, we have used 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 choices (Bailey, Cao, Kuchler, and Stroebel forthcoming; Bailey, Dávila, Kuchler, and Stroebel 2017). We have also explored the role of peer effects in product adoption decisions (Bailey, Kuchler, Stroebel, and Wong 2018), and are working with other coauthors to better understand the role of social connectedness in facilitating social mobility.

For many researchers, it should prove a considerable advantage that the Social Connectedness Index is now more broadly available. In addition to the topics that we have explored in this paper, here are five other examples of policy and research questions that we hope will be pursued with the SCI data.

First, 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 allow researchers and public health officials to better predict where to expect future outbreaks of the flu (Cauchemez et al. 2011; Christakis and Fowler 2010).

Second, the Social Connectedness Index data could also be used to track whether measures of sentiment—for example, those tracked by the Michigan Survey of Consumers or through geo-coded Twitter feeds—spread along social networks.

Third, sociolinguistic research has argued that social networks are an important force determining how languages evolve over time (for example, Milroy 1987). The Social Connectedness Index data would allow researchers to study the extent to which linguistic development in the United States is associated with patterns of social connectedness.

Fourth, the relationships between transportation networks and social connectedness may prove interesting. For example, significant social connectedness between two regions might be a strong indicator that providing transportation infrastructure between these regions, such as direct airline routes, is profitable. Using the Social Connectedness Index as a measure of the potential demand for various routes could address some of the identification issues in the literature analyzing airline scheduling in operations research and industrial organization. Moreover, increased transportation links might also have a causal effect on social connectedness. One approach using the SCI data is to compare the social connectedness 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 connectedness of two similar counties that do not lie on the straight line between major cities (see Bailey et al. 2018).

Finally, the SCI might prove useful in testing theoretical models of network formation (Jackson 2014). Specifically, in models of geographic strategic network formation models, the costs of network formation are directly related to distance (for example, Johnson and Gilles 2000). Using data from the National Longitudinal Survey of Adolescent Health on close friends of individuals, Patacchini, Picard, and Zenou (2015) show that students living in central locations have higher levels of social interactions. Our estimates of the elasticities of friendship links with respect to distance often map directly into the parameters of these models and can be used to parameterize them.

While we hope that the county-level Social Connectedness Index will prove useful to researchers, it is of course only one aspect of the vast wealth of data on networks being created by online social networking services. As these data become available in various forms, the modeling and analysis of social networks will advance substantially.

## References

- Aghion, Philippe, and Peter Howitt.** 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60(2): 323–51.
- 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(2): 258–69.
- Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, Jr.** 2008. "Models of Idea Flows." NBER Working Paper 14135.
- Anderson, James E. and Eric van Wincoop.** 2004. "Trade Costs." *Journal of Economic Literature* 42(3): 691–751.
- Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel.** Forthcoming. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*.
- Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel.** 2017. "House Price Beliefs and Mortgage Leverage Choice." NBER Working Paper 24091.
- Bailey, Michael, Theresa Kuchler, Johannes Stroebel, and Arlene Wong.** 2018. "Peer Effects in

Product Adoption." Unpublished paper.

**Bakshy, Eytan, Dean Eckles, Rong Yan, and Itamar Rosenn.** 2012. "Social Influence in Social Advertising: Evidence from Field Experiments." In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*, pp. 146–61. New York, NY: ACM.

**Bakshy, Eytan, Solomon Messing, and Lada A. Adamic.** 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239): 1130–32.

**Benhabib, Jess, and Mark M. Spiegel.** 2005. "Human Capital and Technology Diffusion." Chap. 13 in *Handbook of Economic Growth*, vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 935–966. Elsevier.

**Berkes, Enrico, and Ruben Gaetani.** 2017. "The Geography of Unconventional Innovation." [https://cpb-us-east-1-juc1ugurlqwqqo4.stackpathdns.com/sites.northwestern.edu/dist/4/638/files/2017/06/Berkes\\_Gaetani\\_Submission\\_June\\_2017-2ao3fck.pdf](https://cpb-us-east-1-juc1ugurlqwqqo4.stackpathdns.com/sites.northwestern.edu/dist/4/638/files/2017/06/Berkes_Gaetani_Submission_June_2017-2ao3fck.pdf).

**Black, Duncan, and Vernon Henderson.** 1999. "A Theory of Urban Growth." *Journal of Political Economy* 107(2): 252–84.

**Blanchard, Olivier Jean, and Lawrence F. Katz.** 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* no. 1, pp. 1–75.

**Buera, Francisco J., and Ezra Oberfield.** 2016. "The Global Diffusion of Ideas." NBER Working Paper 21844.

**Burchardi, Konrad B., and Tarek A. Hassan.** 2013. "The Economic Impact of Social Ties: Evidence from German Reunification." *Quarterly Journal of Economics* 128(3): 1219–71.

**Cauchemez, Simon, Achuyt Bhattarai, Tiffany L. Marchbanks, Ryan P. Fagan, Stephen Ostroff, Neil M. Ferguson, David Swerdlow, and the Pennsylvania H1N1 Working Group.** 2011. "Role of Social Networks in Shaping Disease Transmission during a Community Outbreak of 2009 H1N1 Pandemic Influenza." *PNAS* 108(7): 2825–30.

**Chaney, Thomas.** 2014. "The Network Structure of International Trade." *American Economic Review* 104(11): 3600–34.

**Chaney, Thomas.** 2016. "Networks in International Trade." Chap. 28 in *Oxford Handbook of the Economics of Networks*, edited by Yann Bramoullé, Andrea Galeotti, and Brian Rogers. Oxford University Press.

**Cheng, Justin, Lada A. Adamic, Jon M. Kleinberg, and Jure Leskovec.** 2016. "Do Cascades Recur?" 671–681, International World Wide Web Conferences Steering Committee. arXiv:1602.01107 [cs.SI].

**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." *Quarterly Journal of Economics* 129(4): 1553–1623.

**Chetty, Raj, Michael Stepler, 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* 315(16): 1750–66.

**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." NBER 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., Mikhail Dmitriev, and Esteban Rossi-Hansberg.** 2012. "The Spatial Diffusion of Technology." NBER Working Paper 18534.

**Comin, Diego, and Bart Hobijn.** 2010. "An Exploration of Technology Diffusion." *American Economic Review* 100(5): 2031–59.

**Duggan, Maeve, Nicole B. Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden.** 2015. "Social Media Update 2014." Pew Research Center, January 9.

**Eagle, Nathan, Michael Macy, and Rob Claxton.** 2010. "Network Diversity and Economic Development." *Science*, May 21, 328(5981): 1029–31.

**Facebook.** 2017. "Facebook Form 10-Q, Quarter 4, 2017." [https://s21.q4cdn.com/399680738/files/doc\\_financials/2017/Q4/Q4-2017-Earnings-Presentation.pdf](https://s21.q4cdn.com/399680738/files/doc_financials/2017/Q4/Q4-2017-Earnings-Presentation.pdf).

**Fogli, Alessandra, and Laura Veldkamp.** 2012. "Germs, Social Networks and Growth." NBER Working Paper 18470.

**Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics* 114(3): 739–67.

**Gee, Laura K., Jason Jones, and Moira Burke.** 2017. "Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network." *Journal of Labor Economics* 35(2): 485–518.

**Gee, Laura K., Jason J. Jones, Christopher J. Fariss, Moira Burke, and James H. Fowler.** 2017. "The Paradox of Weak Ties in 55 Countries." *Journal of Economic Behavior & Organization* 133: 362–72.

**Gilbert, Eric, and Karrie Karahalios.** 2009. "Predicting Tie Strength with Social Media." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 211–20. ACM.

**Glaeser, Edward.** 1999. "Learning in Cities."

*Journal of Urban Economics* 46(2): 254–77.

**Granovetter, Mark S.** 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78(6): 1360–80.

**Hampton, Keith, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell.** 2011. “Social Networking Sites and Our Lives.” Pew Internet and American Life Project, Washington, DC.

**Herdagdelen, Amaç, Bogdan State, Lada Adamic, and Winter Mason.** 2016. “The Social Ties of Immigrant Communities in the United States.” *Proceedings of the 8th ACM Conference on Web Science*, 78–84.

**Hobbs, William R., Moira Burke, Nicholas A. Christakis, and James H. Fowler.** 2016. “Online Social Integration is Associated with Reduced Mortality Risk.” *PNAS* 113(46): 12980–984.

**Jackson, Matthew O.** 2014. “Networks in the Understanding of Economic Behaviors.” *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.” *Quarterly Journal of Economics* 108(3): 577–98.

**Johnson, Cathleen, and Robert Gilles.** 2000. “Spatial Social Networks.” *Review of Economic Design* 5(3): 273–99.

**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.” *Review of Economic Studies* 56(4): 569–82.

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

**Lambiotte, Renaud, Vincent D. Blondel, Cristobal 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): 5317–25.

**Lazarsfeld, P., and R. K. Merton.** 1954. “Friendship as a Social Process: A Substantive and Methodological Analysis.” In *Freedom and Control in Modern Society*, edited by Morroe Berger, Theodore Abel, and Charles H. Page. New York: Van Nostrand.

**Liben-Nowell, David, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins.** 2005. “Geographic Routing in Social Networks.” *PNAS* 102(33): 11623–28.

**Lucas, Robert E., Jr.** 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?” *Quarterly Journal of Economics* 121(1): 79–119.

**McPherson, Miller, Lynn Smith-Lovin, and James M. Cook.** 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27: 415–44.

**Milroy, Lesley.** 1987. *Language and Social Networks*. 2nd edition. Wiley-Blackwell.

**Moretti, Enrico.** 1999. “Social Networks and Migrations: Italy 1876–1913.” *International Migration Review* 33(3): 640–57.

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

**Munshi, Kaivan.** 2016. “Community Networks and Migration.” Chap. 23 in *The Oxford Handbook of the Economics of Networks*, edited by Yann Bramoullé, Andrea Galeotti, and Brian Rogers. Oxford University Press.

**Patacchini, Eleonora, Pierre M. Picard, and Yves Zenou.** 2015. “Urban Social Structure, Social Capital and Spatial Proximity.” CEPR Discussion Papers no. DP10501, Center for Economic Policy Research, March.

**Rauch, James E.** 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.” *Journal of Political Economy* 94(5): 1002–37.

**Rupasingha, Anil, Stephan J. Goetz, and David Freshwater.** 2006. “The Production of Social Capital in US Counties.” *Journal of Socio-Economics* 35(1: Essays on Behavioral Economics): 83–101.

**Scellato, Salvatore, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo.** 2011. “Socio-Spatial Properties of Online Location-Based Social Networks.” *Proceedings of the Fifth International Conference on Weblogs and Social Media*, held in Barcelona, Catalonia, Spain, July 17–21. AAAI Digital Library.

**Thompson, Peter, and Melanie Fox-Kean.** 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment.” *American Economic Review* 95(1): 450–60.

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

**Zipf, George Kingsley.** 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.