Distance in Bank Lending: The Role of Social Networks*

Oliver Rehbein†
University of Bonn

Simon Rother‡
University of Bonn

December 31, 2019

Abstract

We analyze the role of social connections in bank lending, using county-to-county friendship-link data from Facebook. Strong social ties increase loan volumes. This effect is distinct from physical distance, which becomes significantly less relevant when accounting for social ties, and from cultural differences, for which we introduce a new measure at the county-pair level. The effect of social connectedness is supply-side driven and particularly large for small banks but demand-side driven for large banks. Less information-sensitive loans are significantly less affected by social connectedness. To bolster identification, we exploit highway connections, historical travel costs, and the quasi-random staggered introduction of Facebook across the US as instruments. Our results reveal the important role of social connectedness as an informal information channel, speak to the nature of borrowing constraints, and have implications for bank-lending and anti-trust policies.

Keywords: bank lending, social networks, information frictions, culture, distance.
JEL-Classification: D82, D83, G21, O16, L14, Z13.

---

*We thank René Bernard (discussant), Martin Brown, Ralph De Haas, Hendrik Hakenes, Luc Laeven, Yuliyan Mitkov, Yueran Ma, Ralf Meisenzahl, Alexandra Niessen-Ruenzi, Farzad Saidi, Martin Schmalz, Isabel Schnabel, Johannes Stroebel and seminar participants in Bonn for valuable comments and suggestions. We also thank participants of the annaul meeting of the German Finance Association in Essen and the workshop on Household Finance in Frankfurt for valuable feedback. Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project C04) and through EXC 2126/1-390838866 under Germany's Excellence Strategy is gratefully acknowledged.

†University of Bonn, Institute for Finance & Statistics, Adenauerallee 24-42, 53113 Bonn, Germany, e-mail: oliver.rehbein@uni-bonn.de

‡University of Bonn, Institute for Finance & Statistics, Adenauerallee 24-42, 53113 Bonn, Germany, e-mail: simon.rother@uni-bonn.de.

1 Introduction

Real-world social networks serve as a channel for the exchange of information and thus can affect economic outcomes, especially in the presence of asymmetric information. During the past decade, these networks have become increasingly visible and wide-spread due to the emergence of online social networks such as Facebook, Twitter, or the Chinese WeChat. This trend is likely to continue as user numbers keep increasing. The information flow through social networks can be particularly relevant if information asymmetries are severe or effort to overcome them is very costly. In bank lending, such costly information frictions play a key role and we analyze whether and how social connectedness influences lending outcomes.

Social connectedness may enhance banks' access to soft information both before and after a loan application. Ex-ante, social connectedness may lead to a higher degree of information about a future loan applicant or, more likely, her local economic environment as a byproduct of social interactions. Such ex-ante information can reduce the need for ex-post information acquisition. It can also reduce the cost of interpreting new borrower-specific information by providing context. Ex post, social connectedness may facilitate the collection of soft information relevant to the specific loan application, because social connections can be leveraged as information source.

We shed light on this new social connectedness channel and provide robust evidence of its relevance for bank lending. The analysis of social connectedness also provides new context to the literature on distance in bank lending. While it is well documented that physical distance poses a significant lending barrier, it remains unclear to what extent this effect is due to transportation costs (Degryse and Ongena, 2005) or due to soft information being local (Agarwal and Hauswald, 2010). We make use of the social connectedness channel to distinguish between these two explanations. Additionally, information flows through social networks would offer a rational for banks' limited ability to collect soft information at large distances. Lastly, the analysis connects to the literature on cultural differences as a lending barrier (Giannetti and Yafeh, 2012; Fisman, Paravisini, and Vig, 2017). These differences can affect lending outcomes by increasing asymmetric information between borrower and lender or due to discrimination. We study in how far social connectedness can compensate for the lending barriers posed by physical and cultural distance as well as the interdependences of the effects of distances and connectedness.

To analyze the social connectedness channel, we use information on social connections from Facebook (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018). The data provide us with the relative probability of a person in county A being friends with a person in county B. We refer to this probability as *social connectedness*. For our analysis, information does not necessarily have to be exchanged via Facebook. Given its 214 million active users in the US, the use of Facebook is fairly wide-spread and corresponds to approximately 65 percent of the population. It is also pervasive across young and old, educated and less educated people and urban and rural areas. Accordingly, the Facebook friendship links relate to real-world social networks in which information can be exchanged offline.

To account for cultural factors, we introduce a new measure which builds on the theoretical models of regional sub-cultures in Elazar (1984) and Lieske (1993). These models consider culture an outcome of a person's ethnic ancestry, religious beliefs, racial origin, and the structure of her social environment. We operationalize these models by collecting data on all four categories at an individual level, which we aggregate at the county-pair level to compute our measure of *cultural distance*.

The lending data covers cross-county lending to small and medium-sized enterprises as well as mortgage lending. Small firms are particularly opaque borrowers such that soft information may play an important role in the credit intermediation process. The mortgage loan data are richer in information and allow us to perform more detailed analyses to determine whether social connectedness actually constitutes an information channel.

Our results reveal that social connectedness matters for county-to-county lending to small and medium-sized enterprises: In our baseline regression, a one standard deviation increase in social connectedness is associated with a 0.17 standard deviations increase in loan volumes. For mortgage lending, the effect is smaller but also significant. The effect of social connectedness is supply-side driven and particularly large for small banks but demand-side driven for large banks. This finding is in line with a less standardized credit intermediation process in small banks, which allows for a more prominent role of soft information. The relevance of the social connectedness channel varies with the information sensitivity of the loan: while it is particularly pronounced for information-sensitive loans, it is less important or entirely muted for loans that are subject to government guarantees or securitization. These results strongly suggest that social connectedness affects bank lending by constituting an informal information channel.

While physical distance has a negative effect on loan volumes, this effect is significantly reduced when accounting for social connectedness. Consequently, relationship factors and soft information can at least partly explain the large distance effects in the literature. While going in opposite directions, the effects of social connectedness and physical distance are similarly large. A one standard deviation increase in social connectedness compensates for approximately 400 miles of additional distance between borrower and lender.

Similarly to physical distance, cultural distance is also associated with lower loan volumes. While cultural proximity and social connectedness both operate through the cost of information, both factors appear to be distinct as they significantly drive bank lending in the same regression. Interestingly, a higher degree of social connectedness reduces the effect of cultural distance. Hence, close social ties help to overcome the negative effects of cultural differences.

Our baseline results rely on cross-sectional regressions using fixed effects to account for source-county and destination-county characteristics. We also control for a number of county-pair specific variables which may influence count-to-county lending, including migration, commuting behavior, and trade. To improve identification, we estimate IV regressions using several distinct instruments. We exploit the structure of the U.S. highway network, the cost of traveling between counties in 1920, and the quasi-random rollout of Facebook in its initial stages as instruments. We find significant first stage correlations between these instruments and the social connectedness between counties today. The IV estimations further corroborate our results on social connectedness. The coefficients of physical distance and cultural distance turn insignificant in some specifications.

The results have a number of implications beyond uncovering the role of social connectedness as informal information channel in bank lending. First, increasing social connectedness over the past decade may provide an additional explanation for the trend towards larger and geographically more dispersed banks. As modern technology facilitates connecting at larger distances, banks can more easily expand across regions. Second, the positive effect of social connectedness on loan volumes suggests that banks and firms can strategically counteract lending barriers posed by large physical or cultural distance by employing teams of well-connected agents. For example, if a bank wishes to expand into Boston, it is probably a good idea to hire personnel with strong social ties in Boston for this expansion. Third, our results imply that regulators might have to take social connections into account in anti-trust decisions. Despite the fact that distance remains an important factor, a high concentration of lenders in a geographical area appears less problematic if banks outside this area are socially well connected with this area.

This paper is closely related to the literature highlighting the importance of social networks in economics. There is a large number of studies in various fields highlighting the importance of (social) networks, for example in product markets, labor markets, trade, stock markets and other financing decisions.[1] Granovetter (2005) highlights that social networks affect the quality of information flows and trust, which are both important in overcoming information asymmetries in lending. In spite of this plausible connection, few papers investigate the role of social connections in the context of banking. Most closely related to our paper is Haselmann, Schoenherr, and Vig (2018), who show that social connections in elite networks leads bankers to extend more (inefficient) credit. While they provide interesting insights into how social connections can affect lending decisions, the elite network they investigate is very different from the broad social network that Facebook provides. Instead of exclusive network membership, we demonstrate that aggregate (county-to-county) social

[1] Jackson (2011) provides an overview about the economic applications of social networks. Knack and Keefer (1997) demonstrates in a cross-country setting that social capital matters for economic growth. Bailey, Johnston, Kuchler, Stroebel, and Wong (2019) study the role of social networks (Facebook) for the adoption of new products. Bailey, Cao, Kuchler, and Stroebel (2018) demonstrate that information about house price developments spreads along socially connected individuals. Ioannides and Datcher-Loury (2004) discuss the role of social networks in labor markets. Nguyen (2012); Kramarz and Thesmar (2013) look at social networks within boards and in upper management of firms. Chaney (2014) investigate the role of networks in international trade. For finance, Duflo and Saez (2003) demonstrate the role of social networks in individual retirement decisions. Several studies demonstrate that both uninformed and informed investors are subject to social signals in their investment decisions (Kelly and Ó Gráda, 2000; Hong, Kubik, and Stein, 2004, 2005; Ivković and Weisbenner, 2007; Brown, Ivković, Smith, and Weisbenner, 2008; Han and Yang, 2013; Halim, Riyanto, and Roy, 2019). Cohen, Frazzini, and Malloy (2008, 2010) demonstrate that mutual fund managers invest more frequently in firms to which they have common social ties, and use this additional information to outperform the market.

connectedness matters for bank lending and importantly, that this social connectedness helps to explain a substantial fraction of the traditional lending distance effect.[2]

More importantly our paper relates to the large literature on relationship banking (Boot, 2000; Kysucky and Norden, 2015) in general and on the effects of physical distance in particular. The effect of physical distance in lending outcomes is highlighted by a long list of influential studies (Petersen and Rajan, 2002; Degryse and Ongena, 2005; Mian, 2006; Agarwal and Hauswald, 2010).[3] Further research indicates that the key factor at play is the collection of soft information, which appears to be more difficult at larger distances (Liberti, 2017).[4] Given the recent advances in information transfer technology, this soft information transfer is perhaps hindered more by differences in social and cultural backgrounds rather than physical transportation cost. Our paper is the first to explicitly tests this idea that the large lending effects of physical distance are better explained by social and cultural factors. This is important, because physical distance has usually been the measure influencing competition policy (Degryse and Ongena, 2005; Granja, Leuz, and Rajan, 2018). However, this paper's result suggest that physical distance is partly a correlate of underlying social and cultural factors, highlighting that competition may occur less on geographical, but more along *socio-cultural* lines, implying that competition policies should take them into account.

[2]There is additional evidence that social connections matter for lending decisions. For example, La Porta, López de Silanes, and Shleifer (2002); Khwaja and Mian (2005) show that political connections matter in lending decisions. Uchida, Udell, and Yamori (2012) show that more frequent interactions with loan officers increases soft information collection from banks, indicating that there is soft information that can only be connected interpersonally. Lin, Prabhala, and Viswanathan (2013) analyze data from a peer-to-peer lending platform and show that lenders' decisions are affected by the behavior of a borrower's online friends. In more sociological research, Uzzi (1999) shows that firms with social ties to banks receive lower interest rates.

[3]Numerous further studies analyze the relationship between physical distance and lending. Berger, Miller, Petersen, Raghuram Rajan, and Stein (2005) shows that larger banks tend to lend at larger distances than smaller banks. Brevoort and Hannan (2006) also demonstrates that physical distance serves as a deterrent to lending activities and that larger banks tend to be less affected by higher distances. DeYoung, Glennon, and Nigro (2008) shows that default probabilities are generally increasing in distance. Beck, Degryse, de Haas, and van Horen (2018) investigate distance in the context of Islamic banks.

[4]Degryse, Laeven, and Ongena (2008) demonstrate that bank's lending distance depends on its organizational structure; less hierarchical banks and banks with better communication tend to lend at shorter distances. (Gropp and Güttler, 2018) demonstrate that smaller German savings banks engage in more collection of soft information. Brei and von Peter (2018) demonstrate that distance is relevant for both lending and trade behavior, even when transportation costs are immaterial, thus highlighting the concern that physical distance may only proxy for informational distance measures.

This paper also contributes to the literature on the importance of culture in banking. Beck, Ioannidou, and Schäfer (2017) highlight that foreign banks have disadvantages in collecting local information, which can be overcome by contract design. Similarly, Houston, Itzkowitz, and Naranjo (2017) demonstrate that such disadvantages can be reduced by owning more foreign assets. Giannetti and Yafeh (2012) demonstrate on a cross country-level that cultural distance drives bank-lending outcomes. Fisman, Paravisini, and Vig (2017) show based on data from India that if loan officer and borrower are similar in terms of caste and religion, more loans tend to be extended and repayment rates increase, suggesting that cultural similarities help to mitigate informational frictions in lending. Our findings are in line with these studies in that loan volumes decrease in the cultural distance between borrower and lender. We contribute to this literature by introducing a measure that quantifies cultural differences between counties in the US and by studying the interplay between cultural distance and social connectedness in the bank-lending context.

2 Data and empirical model

Subsequently, we elaborate on our dataset, its sources, and our empirical model. A concise overview of variable definitions and data sources is provided in Table A1 in the Appendix. In order to understand this section better the reader may find it helpful to keep in mind that our final analysis will be on the county-pair level.

Lending data Our main sample analyzes lending to small and medium-sized enterprises, collected under the Community Reinvestment Act (CRA). In additional analyses, we use lending data from the Home Mortgage Disclosure Act (HMDA)[5]. Both data sets are available through the Federal Financial Institutions Examination Council (FFIEC). The CRA data is aggregated at the bank and destination-county level. It exhibits a broad coverage and comprises loans amounting to over 230 billion USD for 2017.[6] To aggregate each type of lending data at the source-county destination-county level, we assume that each loan is extended by that issuing bank's branch which is closest

[5]For 2017, the HMDA data comprises 14.3 million loans from 5,852 financial institutions.Reporting requirements depend on a number of criteria such as balance sheet size and the number of mortgage loans. These criteria change on a yearly basis. For more information see https://www.ffiec.gov/hmda/default.htm or https://www.consumerfinance.gov/data-research/hmda/learn-more.

[6]For reporting requirements see https://www.ffiec.gov/cra/reporter.htm.

to the destination county.[7] The data on branch locations is provided by the FDIC. Alternatively, we assign origination counties according to banks' headquarter locations. The results are robust.[8]

Social connectedness We use data on social connectedness from Bailey, Cao, Kuchler, Stroebel, and Wong (2018). The measure is based on a cross section of Facebook friendship links from the year 2016, which is aggregated at the county-pair level and scaled by county populations.[9] We winsorize the variable at the 99^{th} percentile and rescale it to the range 0 to 100 for ease of interpretation. Intuitively, the measure captures the relative probability that a person in county A is friends with a person in county B. This probability realistically corresponds to the likelihood of actual cross-county friendships in the US due to the widespread use of Facebook, which is pervasive across young and old, educated and less educated people, or urban and rural areas (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018).

Physical distance We obtain data on the great-circle distance between counties from the National Bureau of Economic Research's (NBER) county distance database. County locations are based on county centroids defined by the US Census Bureau and usually correspond to a county's geographical center. As an alternative, we use the distance between population weighted centroids. In order to capture transportation costs more closely, we additionally use data from the National Transportation Center, which contains distances in highway miles and the costs for the transportation via the cheapest combination of highway, rail, and water ways. The results are robust towards the choice of physical distance measure.[10]

Cultural distance We construct a measure of cultural distance at the US county-pair level based on the theoretical models of regional subcultures in Elazar (1984) and Lieske (1993). These models characterize culture as an outcome of a person's ethnic ancestry, racial origin, religious beliefs, and the structure of her social environment. To operationalize the models, we collect a total of 39

[7] For the HMDA data the destination county is the location of the property securing the mortgage.
[8] See Table A8 for details.
[9] For confidentiality reasons, the variable is also scaled by an unknown scaling factor.
[10] Results using different distance measures are provided in Table 7.

variables on the four categories from the 2010 US Census and the 2010 US Religious Congregations and Membership Study. Table A2 in the appendix provides an overview of the variables in each (sub-)category. For each variable, we calculate the absolute difference per county pair and sum these differences across all variables of one sub-category. Afterwards, we sum across sub-categories to quantify the dissimilarities for each category and, finally, sum across categories to obtain our final measure of cultural distance. To ensure equal contribution to the variation within the final variable, within categories, and within sub-categories, we normalize every summand to mean zero and variance one before calculating the sum.

Further covariates and final dataset We collect the following additional data on the county-pair level: gross migration flows (US Census Bureau), commuting behavior from the source to the destination county (US Census Bureau), state-to-state gross trade volumes (Census Bureau Commodity Flow Survey), and two dummy variables indicating whether county-pairs share a common border (US Census Bureau) and whether they are located in the same state (NBER's county distance database). We also obtain data on counties' three-year average real GDP growth (obtained from the Bureau of Economic Analysis) and unemployment rates (Bureau of Labor Statistics), for which we the absolute value of the county-to-county difference to obtain data at the county-pair level.

This combination of datasets leaves us with 66,684 county pairs for which we have information on cross-county small and medium sized business lending. Table 1 reports summary statistics for the variables used in the analysis. Note that the dataset does not contain within-county loans due to the absence of within-county variation in all other variables. The average volume of loans provided from one county to another county amounts to about 1 million USD, but loan volumes differ significantly across county pairs (standard deviation of 9 million). The median social connectedness between counties is 2 and the median county-to-county distance is about 400 miles.

– Table 1 around here –

Empirical model In our baseline regressions (see Equation 1), we explain loan volumes (in logs) provided from branches in source county i to borrowers in destination county j by their their social connectedness, their physical distance (in logs) and their cultural distance, while controlling for source and destination county fixed effects. We also control for county-pair specific characteristics, which are comprised of the absolute value of the unemployment rate differential, the absolute value of the GDP growth differential, gross migration, trade and commuting flows, and two binary variables indicating whether the source and the destination county are located within the same state or share a common border.

$$\begin{aligned}\ln(\text{loan volume})_{ij} =& \beta_1 \cdot \ln(\text{physical distance})_{ij} + \beta_2 \cdot \text{cultural distance}_{ij} \\ &+ \beta_3 \cdot \text{social connectedness}_{ij} + \gamma \cdot \text{controls}_{ij} + \alpha_i + \alpha_j + \epsilon_{ij}\end{aligned} \tag{1}$$

Loan volumes refer to lending to small and medium-sized enterprises are mortgage lending as indicated in the regression tables. Standard errors are clustered at the source- and destination-county levels.[11] Since the explanatory variables of interest are mostly time-invariant, the regressions rely on cross-sectional data. To mitigate reverse causality concerns, all explanatory variables are lagged by one year. In Section 3.3, we introduce instrumental variable approaches to further rule out potential sources of endogeneity.

3 Results

3.1 Social connectedness and the distance measures

Our analysis focuses on the role of social connectedness as informal information channel that helps to overcome asymmetric information in bank lending. As we also shed light on the role of physical and cultural distances as lending barriers, the subsequent subsection differences and similarities between these factors.

Conceptually, social connectedness, physical and cultural distance are distinct. Abstracting from constituting a proxy for local soft information, physical distance accounts for transportation cost. More specifically, it proxies the actual cost and the opportunity cost of a loan officer driving to

[11] We provide a robustness test with state-level and dyadic clustering in Table A4.

a borrower or vice versa (Degryse and Ongena, 2005). Cultural distance captures the difficulties of interpersonal contact stemming from different sets of believes and backgrounds (Fisman, Paravisini, and Vig, 2017). Social connectedness measures something different; it is the prevalence of friendship networks that can be accessed to acquire information actively and passively. For our analysis, information does not have to be exchanged via Facebook, as the Facebook data serves as a proxy for real world friendship networks that can ease the exchange of information (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018).

Statistically, social connectedness, physical distance, and cultural distance are likely to be correlated. For instance, counties located next to each other should be more socially connected than counties separated by hundreds of miles. In line with the conceptual differences, however, the correlation between the variables is moderate. Physical distance and cultural distance are positively correlated (0.15) while social connectedness exhibits a negative correlation with physical distance (-0.26) and with cultural distance (-0.23) for counties less than 50 miles apart. At larger distances, the correlations are more pronounced (0.43, -0.62, and -0.34), but still far from perfect.

Figure 1 illustrates the differences across the measures by displaying the distances and connectedness between Montgomery County, Ohio and all other counties.[12] Montgomery county is representative for this exercise, as the average correlation between the three variables is at the median of all counties. Dark blue colored areas exhibit the lowest physical distance (Panel (a)), the lowest cultural distance (Panel (b)) and the highest social connectedness (Panel (c)). While most regions, which are socially well-connected to Montgomery County are in or nearby Ohio, there are also significant connections to Florida, Colorado, some counties in North Dakota, and parts of the east cost. The high connectedness to southern Florida is like due to its status as a destination for tourism and retirement out of Ohio. Moreover, Montgomery county's largest employer is the US Air Force. As it turns out, counties with strong social connections to Montgomery county in Idaho and North Dakota all have air force bases. In terms of culture, Ohio exhibits close ties to nearby counties as well, but the geographical distribution of culturally similar counties is much more spread out than the distribution of the social connectedness. While Ohio is also culturally

[12] The largest city in this county is Dayton, Ohio.

close to Florida and parts of the east cost, it shares cultural similarities with counties located in the pacific north-west, to which it is socially less connected.

– Figure 1 around here –

3.2 The role of social connectedness in bank lending

Table 2 displays our baseline regression results of the effect of social connectedness on SME lending. Most importantly, we uncover that a large between-county social connectedness is significantly associated with an increased loan volume (Column (1)). A one-point increase in the social connectedness index is associated with a 1.1% increase in lending. This is a first indication that close social ties may help to overcome information asymmetries in lending. This size is economically meaningful as the standardized beta coefficient at the bottom of the table indicates: a one standard deviation increase in social connectedness is associated with a 0.2 standard deviation increase in loan volumes.

Unsurprisingly, physical distance also significantly affects loan volumes. An increase in physical distance by 1% is associated with a statistically significant decrease in loan volumes by 0.38% (Column (2)). This result at the county-to-county level is in line with the previous literature finding a negative relationship between physical distance and lending activity (Degryse and Ongena, 2005; Agarwal and Hauswald, 2010). The economic size (standardized beta equals -0.24) of this long-established distance effect is just slightly larger than the social connectedness effect from Column (1). When comparing R^2, physical distance also shows a similar explanatory power.[13]

Similarly to social connectedness and physical distance in the first two columns, Column (3) shows statistically significant effects for cultural distance. A one point increase in the cultural distance index is associated with a 3.1% lower loan volume. This finding is in line with a lower degree of asymmetric information between culturally similar borrowers and lenders or by lower costs to overcome these asymmetries. An alternative explanation could be that cultural differences lead to discrimination (statistical or otherwise). The findings based on our new measure of within

[13] In robustness tests, we replace our measure of physical distance with alternative measures of transportation costs between counties. The results are robust (see Table 7).

US county-to-county cultural differences thus supports earlier findings, which were obtained on a cross-country setting (Giannetti and Yafeh, 2012) or based on Indian castes (Fisman, Paravisini, and Vig, 2017).

Column (4) of Table 2 includes all three explanatory variables of interest: social connectedness, physical distance, and cultural distance. The coefficients of these variables all remain highly significant, although all three of them decrease in absolute terms. Comparing the R-squared of the four columns indicates that including all three variables is able to explain the largest share of the variation, indicating again that social connectedness measures something distinct from physical and cultural distance. The importance of the social connectedness channel can also be gleaned from the comparison of the standardized betas. In absolute values, the economic size of the social connectedness channel is almost as large as the physical distance channel, while cultural distance appears to be least important.[14]

– Table 2 around here –

The presented regressions demonstrate a strong association between social connectedness and loan volumes. However, a question remains if accounting for social connectedness matters, or if distance has been a fine proxy of social connectedness. We thus test for differences of the coefficients from the individual regressions (Column (1)-(3)) with the coefficients from the full specification and provide the p-values of this test at the bottom of the table. We find that the coefficient of physical distance decreases significantly when accounting for social connectedness and culture. This provides further evidence that there is a related, yet separate channel which explains the large observed lending decreases over larger distances: social connectedness.

Overall, these baseline findings emphasize the importance of social ties in addition to cultural and physical distance for bank lending outcomes. One potential explanation is that cultural and social connections crucially determine the level of soft information or the costs of acquiring these,

[14]The control variables are either insignificant (GDP growth differences and unemployment) or enter with the expected sign (gross migration, gross trade, same state, common border). The only interesting exception to this is commuting behavior: an increase in the commuting volume from the source to the destination county is associated with less lending volume not more.

thereby helping to overcome information asymmetries. The importance of social connectedness in bank lending is a novel finding and has a number of implications. Banks may be able to improve their ability to collect information and expand their market share by strategically employing a team of well-connected and culturally diverse loan officers. For the same reason that banks may be able to use social connectedness as tools to obtain a greater market share, regulators might want to take social connectedness into account. If bank competition does occur along the dimension of social connectedness in addition to geographical areas, regulators may have to take this factor into account when designing anti-trust policies.

3.3 Instrumental variable approaches

Our previous findings suggest a strong positive association between county-to-county loan volumes and social connectedness. The results are obtained while controlling for source county and destination county fixed effects as well as for a rich set of county-pair level variables, which are lagged by one year. In this section, we propose instrumental variable approaches to deal with potentially remaining endogeneity concerns. Culture and physical distance are slow-moving or not changing at all by their very nature such that they are unlikely or impossible to be driven by bank lending. However, it is conceivable that social connections develop during the lending process or that a third factor drives both variables. While we are controlling for most variables that may be driving both variables, such as migration, trade and commuting, we have not ruled out unobservable factors. We seek to address these endogeneity concerns by using several different instruments to bolster identification. We propose four distinct instruments, which are related to social connectedness but do not directly influence lending outcomes.

As a first instrument for social connectedness, we exploit the current highway system. Counties which share a highway connection should be more socially connected, conditional on the distance between them. We take information on the county-to-county highway connections from Baum-Snow (2007), which also provides us with the historic order in which the highways were constructed. We present the results of this IV regression in Column (1) of Table 3. The first stage correlation is positive and significant. The F-value of around 34 is strong enough to reject concerns regarding a week instrument.[15] Importantly, the social connectedness coefficient remains significant and is slightly larger than in the baseline. However, physical distance and cultural distance turn insignificant. While we remain conservative in assigning too much value to the insignificance of physical distance in this regression, the results may provide a first hint that social connections may be more important than physical distance when it comes to the investigation of lending outcomes.

To reduce concerns that county-to-county lending activity may drive the construction of highways, we next specifically exploit the fact that most highways have been planned and constructed a long time ago. The further back in time highways were constructed, the more time there is for social networks to build, but the less concern there should be that lending volumes today are related to highway construction. We thus exploit a specific feature in the data from Baum-Snow (2007): highway construction dates. From these dates we infer the number of years that the highway has been in existence and use this variable as our second instrument. The results are presented in Column (2) of Table 3. The first stage indicates that longer highway existence is indeed associated with higher social connectedness. The results of the second stage are very similar to the prior regression; again social connectedness remains significant while the distance measures do not.

As an alternative to the two highway instruments, we next make use of historical travel costs. Donaldson and Hornbeck (2016) collects data from the construction period of the railroad in the U.S. to estimate county-to-county travel costs from the turn of the 20th century. The idea behind this instrument is that early travel connections may have long lasting effects in terms of social connectedness until today, whereas they are not causally related to current bank lending. Similar historical identifications have been widely used in the economic literature (e.g., Acemoglu, Johnson, and Robinson, 2001). We show the results based on the historical travel cost instrument in Column (3) of Table 3.[16] The first stage correlation is strong and significant. The result of the second stage is almost the same as in the previous regressions: social connectedness has a positive effect on loan volumes.

– Table 3 around here –

[15] See Table A5 for the full first stage regression results.

[16] Specifically, we use the county-to-county travel cost in 1920.

All three instruments presented so far show very similar results, strongly indicating that social connectedness has a causal effect on county-to-county loan volumes. Nevertheless, we next present a fourth instrument based on a different idea. We exploit the quasi-random staggered introduction of Facebook as an instrument.

In the early days of the social network, Facebook access was rolled out sequentially and in a quasi-random fashion. Created at Harvard by Mark Zuckerberg and colleagues, Facebook memberships were initially restricted to students from Harvard University.[17] Later on, Facebook was gradually opened to other Ivy League colleges and, step by step, to universities across the US. We manually recover the order in which the universities joined Facebook.[18] We combine this order with the universities' location to track how Facebook spread across the US at the county level. To construct this instrument, we rank counties according to this order and construct the following county-pair level instrument from this ranking:[19]

$$Relative\ Facebook\ county\ rank = \frac{Rank_i + Rank_j}{Student\ population_i + Student\ population_j}\ . \tag{2}$$

$Rank_i$ ($Rank_j$) is the rank number of county i (j) in the chronology in which the first member of a university located in that county joined Facebook. The lower the sum of $Rank_i$ and $Rank_j$, the earlier Facebook was present in both counties. We scale this sum by the student populations in 2005, as counties with more students can establish more friendship links.

Importantly, the order in which Facebook spread across counties is independent of today's county-to-county loan volumes. There is a significant time delay between the start of Facebook in 2004 and lending outcomes in 2017. More importantly, however, the order in which Facebook

[17]This restriction was enforced by allowing access only to students with a Harvard University email address.

[18]During the early times of Facebook, members' profile IDs were constructed such that a) students of the same university could be identified based on their user IDs and b) higher user IDs corresponded to universities joining later. Together with publicly available information about the universities early Facebook users visited, this information enables us to manually recover the order in which universities gained access to Facebook.

[19]In some regions, Facebook occurred only after the construction of user IDs had been changed in a way which makes it impossible to recover the date when the first Facebook member in this region created an account. We set the rank of these late joiners to the maximum value of the rank distribution plus one standard deviation

was introduced does not follow any discernible pattern. The spread of Facebook appears to be quasi-random, perhaps with a slight correlation along geographical lines (for which we control).

Column (4) of Table 3 displays the results of using the staggered Facebook roll out as an instrument. Note that we use absolute social connectedness in this regression to avoid a mechanical correlation between the denominator of the instrument (student population) and the denominator of the social connectedness index (scaled by total population).[20] Again this instrument is highly significant in the first stage. Initial Facebook roll out is still relevant for how Facebook friendship links exist today. The second stage results are very similar to the baseline findings, which confirms the original result that social connectedness is highly relevant for lending outcomes.

The results from the IV regressions suggest that the OLS estimates of the effect of social connectedness are - if anything - a lower bound. Using four different instruments to bolster identification, social connectedness remains a highly statistically and economically significant driver of county-to-county loan volumes. At the same time, the statistically significance of physical and cultural distance is reduced in these regressions, highlighting that social connectedness is a central channel influencing loan volumes.

3.4 Bank size, supply, and demand

An important aspect of how social connectedness influences lending volumes lies in the separation of demand and supply: Are more loans demanded if social connectedness is higher or is the demand similar, but loan officers are more inclined to serve this demand? A priori, both channels are plausible. On the one hand, social networks could inform more borrowers about credit offers thus increasing demand. On the other hand, loan officers could leverage social networks to reduce the cost of information acquisition, thus increasing loan supply. We can shed more light on this question by looking at mortgage lending data (HMDA), which includes information on loan applications (i.e., also rejected or withdrawn applications).

At the same time, we hypothesize that social connectedness will play a larger role for smaller banks, especially on the supply side. Given that loan decision processes tend to be more standardized at larger banks, it is less likely that the soft information from social networks can be leveraged

[20]See Column (4) of Table A6 for a robustness test in which we use the original relative social connectedness.

by their loan officer to make decisions on loan applications. Conversely, small banks' loan officers may have more freedom to make loan decision based on information they gather from their social networks.

To analyze these channels, we use the total volume of all filed mortgage loan applications as a measure of loan demand. While it cannot account for borrowers who are discouraged from filing an application in the first place, it should provide a reasonable approximation of loan demand. To identify supply-side effects, we re-estimate our baseline regression but use the volume of loan applications as a demand control, thus leaving us with an indication of how supply reacts to social connectedness. We additionally create a county-pair level variable that accounts for the average county-pair bank-size. Specifically, a "large banks" dummy variable equals one if the average bank size (in terms of total assets) per $ of lending from banks from the source county to the destination county is larger than the median county-to-county bank size. Hence, the variable indicates if on average (per $) banks lending from source to destination county are large. We then interact social connectedness with this large banks indicator to analyze the heterogeneous effects of social connectedness across bank size.

– Table 4 around here –

Column (1) of Table 4 presents the results of that regression for our baseline sample of SME loans. Social connectedness is relevant for all banks, but it appears to matter less for large banks. The coefficients of the regression indicates that the effect of social connectedness is about twice as large for small banks than it is for large banks, which is in line with a more standardized credit intermediation process of large banks leaving little room for soft information. We next use (accepted) "mortgage loans" as the dependent variable in Column (2) with similar outcomes. Social connectedness matters much less for larger banks, with the difference between large and small banks being even more pronounced than for SME lending.

Interestingly, the effect of social connectedness and its dependence on bank size differs across demand and supply. Using mortgage loan applications as a proxy for loan demand, the regressions show that larger banks benefit *more* from social connectedness with regard to loan demand (Column

(3)). This may be due to larger banks ability to advertise over larger regions, potentially leveraging social networks in the process and, consequently, raising demand for their products. The reverse holds when looking at loan supply (Column (4)). Social connectedness is positively associated with larger loan volumes in general, but this effect is much smaller for large banks. Note also that only for loan supply physical distance does not appear to matter once we control for social connectedness, whereas it does matter for loan demand. This is also in line with a transportation costs argument that is largely driven by the demand side.

Overall, Table 4 provides several new and important insights about the role of social connectedness and distance for bank lending. First, in line with expectations, social connectedness matters much more for smaller banks. However, larger banks are the beneficiaries of additional loan demand, whereas smaller banks leverage social connectedness much more for supply decisions. Finally, whereas social connectedness is relevant for supply decisions, physical distance matters only for loan demand. This finding is intuitive, yet novel. Transportation cost matter mostly for demand (the applicant usually drives to the bank and not vice versa), whereas social connections help to overcome information frictions even in supply decisions, especially at small banks. This provides strong evidence that the large effects of physical distance found in the literature may be explained by social connections creating the opportunity to leverage soft information and not actual transportation costs.

3.5 Information sensitivity of loans

The previous results demonstrate that there is a clear link between social connectedness and loan volumes. However, so far the results remain silent on the exact channel through which social connectedness is influencing bank lending. In this section, we analyze more directly whether social connectedness serves as a channel for informal information. To this end, we analyze effects on several types of loans which differ in their information sensitivity.

The results of these regressions are provided in Table 5. We start by explaining mortgage loans which are backed by government guarantees. The HMDA mortgage data identifies such loans, most of which are secured by the Federal Housing Administration or the Veterans Administration. We aggregate county-to-county mortgage lending with and without guarantees separately and run our baseline regression on these two different subsets of mortgage lending. Column (1) displays the results for loans which are not backed by government guarantees. For these loans, social connectedness is highly significant and the effect is almost twice as large as for all loans. This suggests that social networks play an increasingly important role if the bank is taking on the full risk of the loan, suggesting strongly that they are valuing the social information they receive more highly. This interpretation is supported by the results from Column (2), which show that social connectedness is not important if the loans are backed by government guarantees. This is consistent with the fact that banks exploit social connections as a tool to overcome information asymmetries, especially if that information is relevant for them; if they can pass on the risk to the government, they do not need to make use of this information. We test formally if there is a difference between government guaranteed loans and ordinary loans in Column (3), where we interact social connectedness with the volume of guaranteed loans in a regression on overall loans. As can be seen, having more guaranteed loans significantly decreases the importance of the social connectedness channel.

– Table 5 around here –

Furthermore, the HMDA data allows us to differentiate between loans where the full risk of the loan remains on the banks balance sheet and loans which are sold to other institutions such as Fannie Mae or Freddie Mac. If the effect of social connectedness runs through an information channel, the effect of social connectedness should be less strong for the less information-sensitive securitized loans. This is in fact confirmed by the results which are displayed in Column (4)-(6). The social connectedness coefficient is about twice as large for loans which are not sold, compared to loans where the bank is not taking on the full risk of the loan, although even sold loans still show higher volumes from increased social connectedness. This is again confirmed in interactions: the more loans are sold to other banks, the less important the social connectedness channel becomes for loan volumes.

These results strongly support the idea of banks using social soft information actively and on purpose if they have the incentives to do so. Once the bank bears the full risk of the loan, it actively appears to use soft information collected through social channels to increase their lending activity, whereas it does not do so in case of less information-sensitive loans.

3.6 Interdependence of effects

So far we have considered only linear effects of physical distance, cultural distance and social connectedness. However, it may be useful to also consider non-linear effects and interactions between the three. In particular there are three interesting questions. Is the effect of social connectedness on loans non-linear? Does social connectedness change the effect of physical distance? Does the effect of cultural distance depend on social connectedness?

We analyze these questions making use of interactions terms. The results are displayed in Table 6. We begin by investigating whether the effect of social connectedness follows a non-linear path by including the social connectedness variable in its squared form in Column (2). For comparison, we restate the baseline in Column(1). We find that the squared term of social connectedness is statistically insignificant. This indicates that there are no very clear non-linearities of the effect of social connectedness on loan volumes. Note that while even the non-squared term turns insignificant, the coefficients are jointly significant, such that this regression is not evidence of an absence of an overall social connectedness effect in our regression.

Do social connections become more or less important for loan volumes as distance increases? The results in Column (3) suggest the latter, as the interaction between physical distance and social connected is negative and significant. This implies that as distance increases, the effect of social connectedness on loan volumes is decreasing. We surmise that this decreasing effect may be due to the strength of social networks at larger distances. Often if distances are large, relationships become less intense as opportunities for face-to-face contact fade. As a result it may be more difficult to leverage social connections for loan decisions (supply and demand) and as a result social connections may be less important for loan volumes at larger distances.

– Table 6 around here –

While social connections cannot well compensate for physical distances, Column (4) suggests that it can indeed compensate for larger *cultural distances*. The interaction term between social connectedness and cultural distance is significant and positive. If cultural distance is larger, social connectedness becomes more important. This is a very interesting and intuitive result: social networks can help to bridge the cultural divide, at least when it comes to making decisions on loans. Note that there are two potential explanations for this effect: social connectedness can help to overcome differences that are based in statistical differences between cultures through enabling more differentiated information flows (limiting statistical discrimination) or perhaps social connectedness lowers personal discriminatory practices as it tears down barriers stemming from (subconscious) prejudices that result in discrimination absent close personal connections.

3.7 Robustness

In this section we address some potential concerns that might remain after the main analysis. We first mute concerns that the physical distance measure of county-to-county is biasing our results, by providing estimations using several different distance measures in Table 7. For ease of comparison, we restate the baseline, which uses county-centroid to county-centroid Great Circle Distance ("as the crow flies") as the measure of physical distance. However, there might be concern that if the population is not located on average near the center of the county, we may underestimate physical distance, potentially making our measure of social connectedness pick up this variation that only stems from inaccuracies in controlling for this concern. To address this potential issue, Column (2) uses population weighted centroids to calculate county to county physical distance and the results are almost exactly the same as in the baseline.

– Table 7 around here –

We then move to the concern that Great Circle distance may not properly capture real transportation costs, as those depend not only on the air distance but also on how well connected counties are through highways and other modes of transportation. To this end, we use data from the Oak Ridge National Laboraty's National Transportation Center to gage the actual transportation costs between county pairs. Column (3) uses highway travel costs and Column (4) includes the cost measure using all modes of transportation. The table indicates no change from the baseline of the social connectedness coefficient and differences to the standard physical distance measure from the baseline is marginal.

While the focus of this paper is the effect of social connectedness on loan volumes, we also introduce a novel measure of within-county cultural distance which is statistically significant and economically relevant in many of our specification in predicting loan volumes. Because of the novelty of this measure we now undertake a few steps to demonstrate that it measures something meaningful. We first conduct a cluster analysis with our measure of cultural distance. To this end, we first select the 10 principal components of the cultural variable, and then sort counties into seven clusters by minimizing the euclidean distances between the principal components' scores. We then display these seven clusters of cultural distance that result from this data-driven cluster selection in Figure A1. A visual inspection demonstrates that these cultural clusters based on 39 individual variables provide a pretty nice picture of well-known regional sub-cultures in the US.[21]

– Table 8 around here –

Despite the fact that the cluster analysis suggests that our new culture variable does not measure something completely intangible, we intend to mute concerns that the construction of the variable itself is driving our results on either social connectedness or culture. To this end we first include our cultural distance measure in sub-categories in Column (2) and (3). The results indicate that mainly the socio-economic differences are driving the cultural differences in the data, although race plays a slight role, once one does not account for the socio-economic differences of the county-pairs. Column (4) finally includes all 39 culture variables individually (coefficients not reported). We do this to ensure that the effect of social connectedness does not depend on the exact construction of the culture variable, which is clearly the case. The effect of social connectedness of loan volumes remains exactly the same.

Lastly, we want to test if other ways of measuring cultural differences yield similar results. One indicator of cultural differences is voting behavior. While this is more an outcome indicator

[21]Note that the cluster analysis appears to suggest one cultural region for large cities, independent from their geographical location.

of cultural difference rather than a fundamental way to estimate it, it is nevertheless interesting to see if the intuition works in a similar way. Consequently, we download county-level data on the controversial 2016 election. We calculate the difference of the share of votes for the republican candidate between counties and use this difference as a regressor in Column (5). The results are in line with our previous results. An increase in the difference of the share of republican votes by 1 percentage point is associated with a decrease in county-to-county loan volumes by 0.6%. This is in line with our baseline results that cultural differences are - in addition to social connectedness - highly relevant for lending outcomes even in a within-country setting. Importantly, even when using a drastically different measure to control for cultural distance, the social connectedness coefficient remains unchanged, highlighting the result that social connectedness is highly important in banking.

4 Conclusion

We find that social connectedness matters for loan volumes providing at least a partial explanation of the very large effects on distance on lending outcomes in the prior literature. Using county-to-county loan data for both small business lending and mortgage lending combined with data on social connections from Facebook and a newly constructed measure of cultural distance, our analysis offers several important insights.

First, social connections appear to be important in overcoming information frictions in bank lending. Stronger social ties between two counties lead to significantly higher loan volumes. Our results indicate that this effect is particularly important for small banks, more important for small business lending, and matters both for loan demand and loan supply. Smaller banks benefit more from social connectedness on the supply side, while large banks benefit from social connectedness through the demand side.

Second, we demonstrate that cultural differences matter for bank lending outcomes as a larger cultural distance between counties is associated with lower loan volumes. While previous studies have found similar effects, our results highlight that culture plays a role for loan volumes even within the US. The results thus demonstrate that cultural differences matter even within a single country, where culture is relatively homogeneous. Moreover, we find evidence that the negative effects of cultural differences disappear if counties exhibit a sufficiently large social connectedness.

Our results strongly suggest that the distance between lender and borrower may need to be increasingly understood as social and cultural distance and less as physical distance. This interpretation of the distance effect thus provides a solution to the puzzle for the economically large effects in the literature on the role of physical distance in bank lending, which seem to be at odds with the advances in information technologies and the seemingly small role of transportation costs during the lending process. Depending on the specification, our results suggest that the strong effects of physical distance on loan volumes can disappear entirely, once controlling properly for social connectedness and cultural differences, especially for supply decisions.

These results also have a number of concrete implications, for example with regard to the design of antitrust policies, for which it appears advisable to consider cultural differences and the social ties as these factors crucially determine bank lending outcomes. Perhaps most concretely our results can be strategically used by banks in the lending process. Banks can benefit from social connectedness to external counties, if they employ loan officers and other decisions maker which are socially connected to such counties. As a result, banks may want to pay attention to their employees social networks when making hiring decisions and can strategically make use of the social connections to outside regions.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

Agarwal, S., and R. Hauswald (2010): "Distance and private information in lending," *The Review of Financial Studies*, 23(7), 2757–2788.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018): "The economic effects of social networks: Evidence from the housing market," *Journal of Political Economy*, 126(6), 2224–2276.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018): "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 32(3), 259–280.

Bailey, M., D. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2019): "Peer effects in product adoption," *National Bureau of Economic Research: Working Paper*.

Baum-Snow, N. (2007): "Did Highways Cause Suburbanization?," *Quarterly Journal of Economics*, 122(2), 775–805.

Beck, T., H. Degryse, R. de Haas, and N. van Horen (2018): "When arm's length is too far: Relationship banking over the credit cycle," *Journal of Financial Economics*, 127(1), 174–196.

Beck, T., V. Ioannidou, and L. Schäfer (2017): "Foreigners vs. natives: Bank lending technologies and loan pricing," *Management Science*.

Berger, A., N. Miller, M. Petersen, Raghuram Rajan, and J. Stein (2005): "Does function follow organizational form? Evidence from the lending practices of large and small banks," *Journal of Financial Economics*, 76(2), 237–269.

Boot, A. (2000): "Relationship Banking: What Do We Know?," *Journal of Financial Intermediation*.

Brei, M., and G. von Peter (2018): "The distance effect in banking and trade," *Journal of International Money and Finance*, 81, 116–137.

Brevoort, K. P., and T. H. Hannan (2006): "Commercial lending and distance: evidence from Community Reinvestment Act data," *Journal of Money, Credit and Banking*, pp. 1991–2012.

Brown, J. R., Z. Ivković, P. A. Smith, and S. J. Weisbenner (2008): "Neighbors matter: Causal community effects and stock market participation," *The Journal of Finance*, 63(3), 1509–1531.

Cameron, C. A., and D. L. Miller (2014): "Robust Inference for Dyadic Data," *Working Paper*.

Chaney, T. (2014): "The network structure of international trade," *The American Economic Review*, 104(11), 3600–3634.

Cohen, L., A. Frazzini, and C. Malloy (2008): "The small world of investing: Board connections and mutual fund returns," *Journal of Political Economy*, 116(5), 951–979.

——— (2010): "Sell-side school ties," *The Journal of Finance*, 65(4), 1409–1438.

Degryse, H., L. Laeven, and S. Ongena (2008): "The impact of organizational structure and lending technology on banking competition," *Review of Finance*, 13(2), 225–259.

Degryse, H., and S. Ongena (2005): "Distance, lending relationships, and competition," *The Journal of Finance*, 60(1), 231–266.

DeYoung, R., D. Glennon, and P. Nigro (2008): "Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers," *Journal of Financial Intermediation*, 17(1), 113–143.

Donaldson, D., and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 131(2), 799–858.

Duflo, E., and E. Saez (2003): "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, 118(3), 815–842.

Elazar, D. J. (1984): *American Federalism.* Harper & Row, New York, 3 edn.

Fisman, R., D. Paravisini, and V. Vig (2017): "Cultural proximity and loan outcomes," *The American Economic Review*, 107(2), 457–492.

Giannetti, M., and Y. Yafeh (2012): "Do cultural differences between contracting parties matter? Evidence from syndicated bank loans," *Management Science*, 58(2), 365–383.

Granja, J., C. Leuz, and R. G. Rajan (2018): "Going the extra mile: Distant lending and credit cycles," *National Bureau of Economic Research: Working Paper*, No. 25196.

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

Gropp, R., and A. Güttler (2018): "Hidden gems and borrowers with dirty little secrets: Investment in soft information, borrower self-selection and competition," *Journal of Banking & Finance*, 87 (2018), 26–39.

Halim, E., Y. Riyanto, and N. Roy (2019): "Costly Information Acquisition, Social Networks, and Asset Prices: Experimental Evidence," *The Journal of Finance*.

Han, B., and L. Yang (2013): "Social networks, information acquisition, and asset prices," *Management Science*, 59(6), 1444–1457.

Haselmann, R., D. Schoenherr, and V. Vig (2018): "Rent seeking in elite networks," *Journal of Political Economy*, 126(4), 1638–1690.

Hong, H. G., J. D. Kubik, and J. C. Stein (2004): "Social interaction and stock-market participation," *The Journal of Finance*, 59(1), 137–163.

——— (2005): "Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers," *The Journal of Finance*, 60(6), 2801–2824.

Houston, J. F., J. Itzkowitz, and A. Naranjo (2017): "Borrowing beyond borders: Foreign assets, lender choice, and loan pricing in the syndicated bank loan market," *Journal of Corporate Finance*, 42, 315–334.

Ioannides, Y. M., and L. P. Datcher-Loury (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42(4), 1056–1093.

Ivković, Z., and S. J. Weisbenner (2007): "Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices," *The Review of Financial Studies*, 20(4), 1327–1357.

Jackson, M. O. (2011): "An overview of social networks and economic applications," in *Handbook of social economics*, pp. 511–585. Elsevier, North-Holland, Amsterdam [u.a.].

Kelly, M., and C. Ó Gráda (2000): "Market contagion: Evidence from the panics of 1854 and 1857," *The American Economic Review*, 90(5), 1110–1124.

Khwaja, A. I., and A. Mian (2005): "Do lenders favor politically connected firms? Rent provision in an emerging financial market," *The Quarterly Journal of Economics*, 120(4), 1371–1411.

Knack, S., and P. Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *The Quarterly Journal of Economics*, 112(4), 1251–1288.

Kramarz, F., and D. Thesmar (2013): "Social networks in the boardroom," *Journal of the European Economic Association*, 11(4), 780–807.

Kysucky, V., and L. Norden (2015): "The benefits of relationship lending in a cross-country context: A meta-analysis," *Management Science*, 62(1), 90–110.

La Porta, R., F. López de Silanes, and A. Shleifer (2002): "Government ownership of banks," *The Journal of Finance*, 57(1), 265–301.

Liberti, J. M. (2017): "Initiative, incentives, and soft information," *Management Science*.

Lieske, J. (1993): “Regional subcultures of the United States,” *The Journal of Politics*, 55(4), 888–913.

Lin, M., N. R. Prabhala, and S. Viswanathan (2013): “Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending,” *Management Science*, 59(1), 17–35.

Mian, A. (2006): “Distance constraints: The limits of foreign lending in poor economies,” *The Journal of Finance*, 61(3), 1465–1505.

Nguyen, B. D. (2012): “Does the rolodex matter? Corporate elite’s small world and the effectiveness of boards of directors,” *Management Science*, 58(2), 236–252.

Petersen, M. A., and R. G. Rajan (2002): “Does distance still matter? The information revolution in small business lending,” *The Journal of Finance*, 57(6), 2533–2570.

Uchida, H., G. F. Udell, and N. Yamori (2012): “Loan officers and relationship lending to SMEs,” *Journal of Financial Intermediation*, 21(1), 97–122.

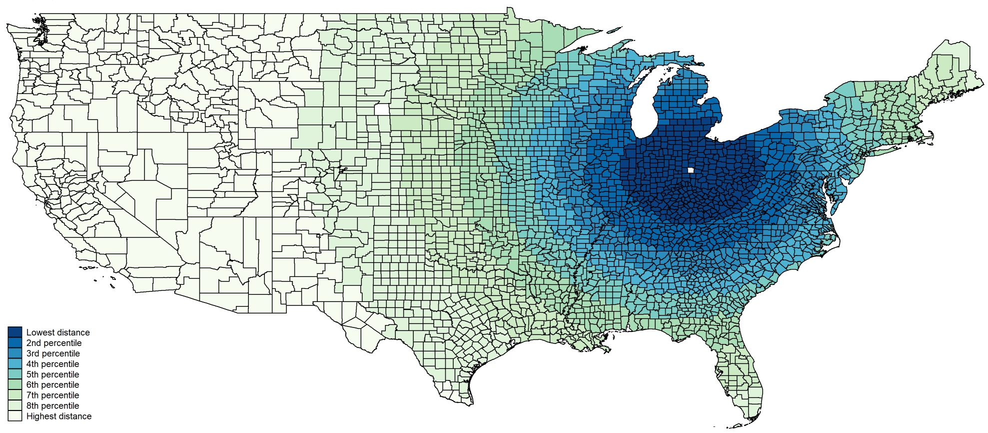
Uzzi, B. (1999): “Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing,” *American Sociological Review*, 64(4), 481.

Figures and Tables

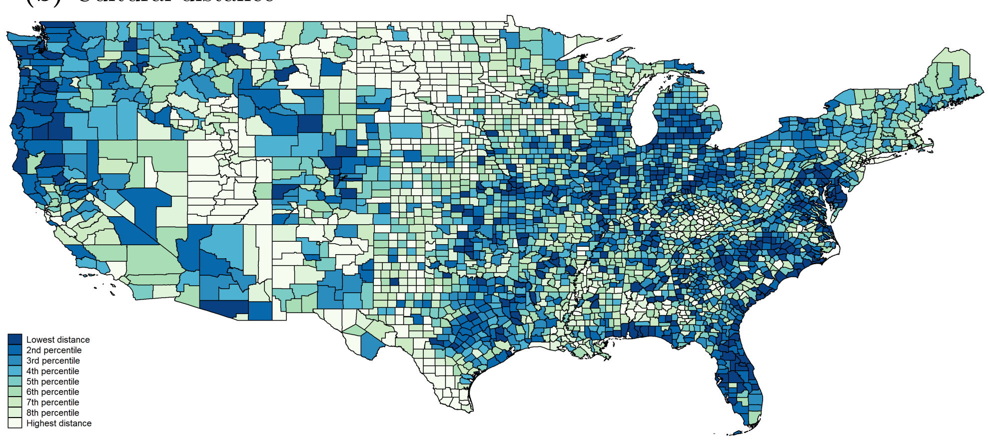
Figure 1: Distances and connectedness between Montgomery County, Ohio and all other counties

The Figure displays (a) *physical distance*, (b) *cultural distance*, and (c) *social connectedness* between Montgomery County, Ohio and all other US counties. This county is representative as the average correlation between the three variables is at the median of all counties. For comparability, dark blue colored areas exhibit the lowest physical and cultural distance, *but the highest* social connectedness.

(a) Physical distance



(b) Cultural distance



(c) Social connectedness

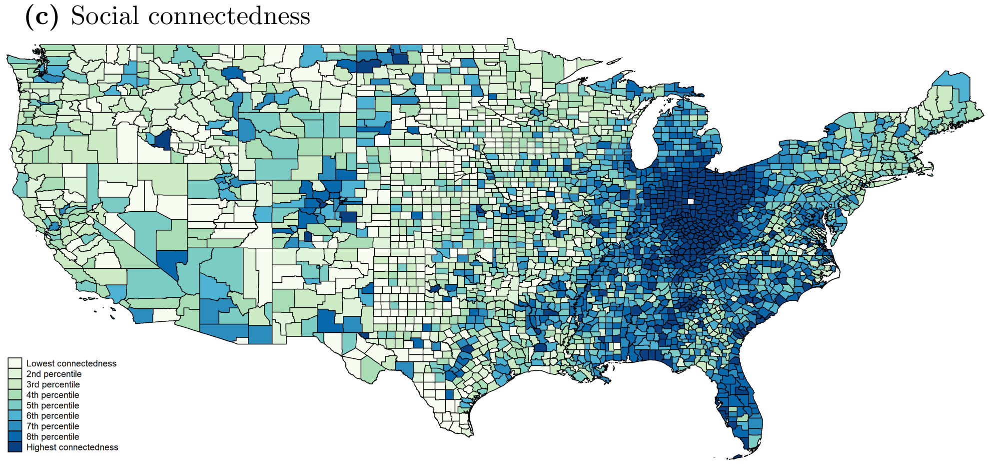


Table 1: Descriptive statistics at county-pair level

The table displays statistics for the variables used in the main analyses. All variables are at the county-pair level. For instance, "volume of SME loans" captures the volume of loans provided by all bank branches in a source county to all small and medium-sized enterprises in the destination county. Variable definitions and data sources are summarized in Table A1 in the appendix. Table A3 provides descriptive statistics for the mortgage loan sample used for the distinction between supply and demand-side effects (see Table 4, columns 2 to 4).

	N	Mean	Median	SD	Min	Max
Volume of SME loans [thousand USD]	66,684	1,057	142	9,132	0	1,296,303
log(Volume of SME loans)	66,684	11.9	11.9	2.0	0.0	21.0
Social connectedness	66,684	20	2	35	0	100
Physical distance [miles]	66,684	578	413	566	4	4,996
log(Physcial distance)	66,684	5.8	6.0	1.3	1.5	8.5
Cultural distance	66,684	17	16	7	0	72
GDP growth differential	66,684	3.6	2.7	3.4	0.0	46.9
Unemployment differential	66,684	1.5	1.1	1.4	0.0	21.4
Gross migration	66,684	108	0	346	0	2,327
Gross trade [million USD]	66,684	85	38	114	0	814
% commuting	66,684	0.2	0.0	0.9	0.0	28.4
Same state	66,684	0.2	0	0.4	0	1
Common border	66,684	0.1	0	0.3	0	1
Same highway	66,684	0.1	0	0.3	0	1
Years since highway construction	66,684	5	0	15	0	79
Historical travel costs	56,265	7	5	4	1	38
Relative Facebook county rank	57,105	0.04	0.02	0.05	0.00	0.48

Table 2: The role of distance and connectedness

The dependent variable is the logarithm of the volume of cross-county loans to small and medium-sized enterprises. The bottom part of the table informs about the statistical significance of the difference of coefficients in columns 1 to 3 and in column 4. The standardized beta coefficients at the end of the table express the effect of a one standard deviation increase in the explanatory variable in standard deviations of the dependent variable. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dependent variable:	log(Volume of SME loans)			
Social connectedness	0.011***			0.007***
	(0.001)			(0.001)
Physical distance		-0.376***		-0.263***
		(0.040)		(0.048)
Cultural distance			-0.031***	-0.014*
			(0.007)	(0.007)
GDP growth differential	-0.008	-0.007	-0.006	-0.006
	(0.007)	(0.008)	(0.008)	(0.008)
Unemployment differential	-0.016	-0.020	-0.010	-0.004
	(0.021)	(0.022)	(0.020)	(0.020)
Gross migration	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Gross trade	0.001***	0.000	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
% commuting	-0.072***	-0.069***	-0.060***	-0.064***
	(0.011)	(0.010)	(0.010)	(0.010)
Same state	0.305***	0.274***	0.716***	0.092
	(0.067)	(0.084)	(0.069)	(0.068)
Common border	0.925***	0.963***	1.201***	0.821***
	(0.052)	(0.080)	(0.058)	(0.063)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684	66,684
Adj. R^2	0.520	0.522	0.516	0.525
Adj. R^2 within	0.123	0.127	0.116	0.132
P-value for H0: no difference to coefficient in column (4)				
Social connectedness	0.002			
Physical distance		0.066		
Cultural distance			0.097	
Standardized beta coefficients				
Social connectedness	0.2			0.12
Physical distance		-0.24		-0.17
Cultural distance			-0.11	-0.05

Table 3: The role of distance and connectedness: instrumental variable approaches

The dependent variable is the logarithm of the volume of cross-county loans to small and medium-sized enterprises. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. First-stage coefficients, p-values, and F-statistics of the instruments are reported in the bottom section of the table. Full first stage regressions are reported in Table A5. In column 4, we use an *absolute* measure of social connectedness while additionally controlling for the inverse population product of the county pair to exclude a mechanical correlation between the *relative* county rank and the *relative* social connectedness. The results are robust to using relative social connectedness (see Table A6, column 1). Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dependent variable:		log(volume of SME loans)		
Instrument:	same highway	years since construction	historical travel cost	facebook rollout
Social connectedness	0.027**	0.027*	0.027***	0.012**
	(0.013)	(0.014)	(0.005)	(0.006)
Physical distance	-0.038	-0.042	-0.031	-0.168
	(0.153)	(0.161)	(0.073)	(0.126)
Cultural distance	-0.005	-0.006	-0.005	0.009
	(0.010)	(0.010)	(0.007)	(0.009)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	56,223	56,852
Adj. R^2	0.507	0.507	0.515	0.541
Adj. R^2 within	0.098	0.100	0.102	0.116
Instrument (1st stage)	2.951***	0.055***	4.07***	161.506***
	(0.000)	(0.000)	(0.000)	(0.000)
F-value (1st stage)	34.341	30.52	136.6	163.5

Table 4: Heterogeneities across bank size, supply, and demand

"Large banks" is a dummy variable which equals one if the average bank size per USD of loans provided from a bank in the source county to a customer in the destination county is larger than the median bank size per USD of cross-county loans. This variable is also included as a single term in all four columns. Additional control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dep. var.: log(Volume of ...)	SME loans	Mortgage loans	Mortgage loan demand	Mortgage loan supply
Social connectedness	0.010***	0.027***	0.003***	0.024***
	(0.001)	(0.003)	(0.001)	(0.003)
Social connectedness · Large banks	-0.006***	-0.039***	0.003***	-0.042***
	(0.001)	(0.003)	(0.001)	(0.003)
Physical distance	-0.184***	-0.030	-0.090***	0.087
	(0.043)	(0.111)	(0.034)	(0.095)
Cultural distance	-0.015**	-0.015	-0.004	-0.010
	(0.007)	(0.012)	(0.004)	(0.011)
County-pair level controls	Yes	Yes	Yes	Yes
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
Loan demand control	No	No	No	Yes
No. of obs.	66,684	34,483	34,483	34,483
Adj. R^2	0.538	0.502	0.412	0.588
Adj. R^2 within	0.155	0.439	0.198	0.536

Table 5: Heterogeneities across the information sensitivity of loans

Columns 1, 2, 4, and 5 use sample splits based on government guarantees and securitization. The number of observations is identical to the full mortgage loan sample as we perform the sample split on the loan level, whereas the regressions are estimated on the county-pair level. Columns 3 and 6 use interaction terms instead of sample splits. "Loans with guarantees" refers to the volume of mortgage loans subject to government guarantees and is also included as a single term in column 3. "Loans sold" refers to the volume of mortgage loans sold off book and is included as a single term in column 6. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. var.: log(Volume of mortage loans ...)	(1) without guarantees	(2) with guarantees	(3) of both types	(4) kept on book	(5) sold off book	(6) of both types
Social connectedness	0.016***	0.004	0.015***	0.013***	0.007**	0.027***
	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)	(0.003)
Social connectedness · Loans with guarantees			-0.002***			
			(0.000)			
Social connectedness · Loans sold						-0.003***
						(0.000)
Physical distance	-0.405***	-0.864***	-0.115	-0.421***	-0.553***	-0.162*
	(0.152)	(0.151)	(0.098)	(0.145)	(0.144)	(0.083)
Cultural distance	-0.045***	-0.018	-0.018*	-0.002	-0.019	-0.008
	(0.017)	(0.021)	(0.010)	(0.013)	(0.016)	(0.007)
County-pair level controls	Yes	Yes	Yes	Yes	Yes	Yes
Source county FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	34,483	34,483	34,483	34,483	34,483	34,483
Adj. R^2	0.262	0.297	0.342	0.365	0.318	0.588
Adj. R^2 within	0.066	0.044	0.259	0.105	0.045	0.536

Table 6: Non-linear effects of social connectedness

Column 1 restates the baseline regression from Table 2, column 4. Columns 2 to 4 add a squared term of social connectedness or interactions between social connectedness and the two distance measures to our baseline regression. The standardized beta coefficients at the end of the table express the effect of a one standard deviation increase in the explanatory variable in standard deviations of the dependent variable. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dependent variable:	log(Volume of SME loans)			
Social connectedness	0.0068***	0.0051	0.0266***	0.0027**
	(0.001)	(0.003)	(0.003)	(0.001)
Social connectedness^2		0.0000		
		(0.000)		
Social connectedness · Physical distance			-0.0049***	
			(0.001)	
Social connectedness · Cultural distance				0.0004***
				(0.000)
Physical distance	-0.2629***	-0.2696***	-0.2425***	-0.2443***
	(0.048)	(0.053)	(0.048)	(0.050)
Cultural distance	-0.0139*	-0.0141**	-0.0159**	-0.0193**
	(0.007)	(0.007)	(0.007)	(0.008)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684	66,684
Adj. R^2	0.525	0.525	0.526	0.526
Adj. R^2 within	0.132	0.132	0.135	0.134

Table 7: Alternative measures of physical distance

Column 1 restates the baseline regression from Table 2, column 4. Columns 2 to 4 re-estimate the baseline regressions with alternative measures of physical distance. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dependent variable:	log(Volume of SME loans)			
Social connectedness	0.007***	0.007***	0.007***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Physical distance	-0.263***			
	(0.000)			
Physical distance (population-weighted centroids)		-0.259***		
		(0.000)		
Highway travel cost			-0.260***	
			(0.000)	
All modes travel cost				-0.264***
				(0.000)
Cultural distance	-0.014*	-0.014*	-0.015**	-0.016**
	(0.055)	(0.053)	(0.040)	(0.025)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
No. of obs.	66,647	66,647	66,647	66,647
Adj. R^2	0.525	0.525	0.524	0.524
Adj. R^2 within	0.132	0.132	0.131	0.131

Table 8: Alternative specifications of cultural distance

Column 1 restates the baseline regression from Table 2, column 4. Columns 2 to 4 re-estimate the baseline regressions with alternative specifications of cultural distance. Column 5 uses the absolute difference of the vote share for the republican candidate in the 2016 presindential election to proxy for culture. Control variables are GDP growth and unemployment differentials, gross migration and trade, as well as same state and common border indicator variables. "Individual culture controls" refers to the 39 variables used to construct our measure of cultural distance (see Table A2). Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	log(Volume of SME loans)				
Social connectedness	0.007***	0.007***	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Physical distance	-0.263***	-0.279***	-0.271***	-0.310***	-0.287***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cultural distance	-0.014*				
	(0.056)				
Culture: ethnic ancestry		-0.011	-0.027		
		(0.803)	(0.522)		
Culture: racial origin		-0.011	-0.075*		
		(0.836)	(0.095)		
Culture: religious beliefs		-0.000	-0.000		
		(0.996)	(0.995)		
Culture: social environment		-0.113**			
		(0.022)			
Vote-share differential					-0.660**
					(0.032)
Source county FE	Yes	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Individual culture controls	No	No	No	Yes	No
No. of obs.	66,684	66,684	66,684	66,684	66,377
Adj. R^2	0.525	0.525	0.525	0.531	0.525
Adj. R^2 within	0.132	0.133	0.132	0.144	0.134

Appendix

A1 Additional Figures and Tables

Figure A1: Clusters of regional subcultures

The figure displays regional subcultures in the US based on a cluster analysis. For this analysis, we determine the principal components of our 39 individual culture variables. We keep the 10 principal components with Eigenvalues larger than one. Afterwards, an algorithm sorts counties into seven clusters by minimizing the mean of the Euclidean distances between the principal components' scores. The number of clusters is chosen based on an elbow method with 1000 repetitions of randomly chosen starting counties for each cluster. The figure's pattern is robust to the number of principal components and clusters.

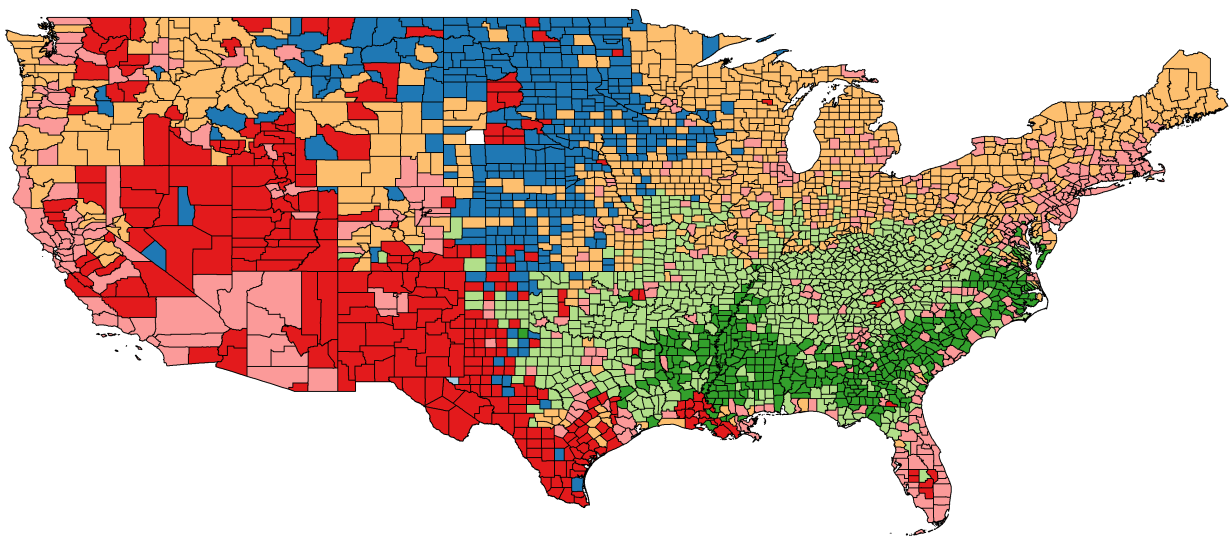


Table A1: Variable definitions and data sources

This table presents provides an overview of variable definitions and data sources. Section 2 contains detailed variable descriptions. Summary statistics are displayed in Tables 1 and A3.

Variable name	**Description**
Dependent variables	
Volume of SME loans	Volume of newly originated loans to small and medium-sized enterprises from source to destination county; enters regressions in logs; source: CRA.
Volume of mortgage loans	Volume of newly originated mortgage loans from source to destination county; enters regressions in logs; source: HMDA.
Mortgage loan demand	Volume of mortgage loan applications from destination county to source county; enters regressions in logs; source: HMDA.
Main explanatory variables	
Social connectedness	Relative probability of friendship links between source and destination county; scaled to [0;100]; source: Bailey, Cao, Kuchler, Stroebel, and Wong (2018).
Physcial distance	Great circle distance in miles based on county centroids; source: NBER's county distance database.
Cultural distance	Index quantifying the cultural distance between two counties; scaled to [0,100]; calculation described in Section 2; source: own calculation.
Instruments	
Same highway	Binary variable equal one if source and destination county are connected by the same highway; source: Baum-Snow (2007).
Years since construction	Number of years since the end of the construction of the highway connecting source and destination county; source: Baum-Snow (2007).
Historical travel cost	Costs of traveling from source to destination county in 1920; source: Donaldson and Hornbeck (2016).
Relative Facebook county rank	The sum of the order (rank) in which Facebook became available in the source county and the destination county, divided by the sum of the student population in both counties (see Section 3.3 and Equation 2). The staggered introduction of Facebook across the US was quasi-random but still explains Facebook friendship patterns today.
Main control variables	
GDP growth differential	Absolute value of the county-pair difference in the average real GDP growth during the last 3 years; in percentage points; source: US Bureau of Economic Analysis.
Unemployment differential	Absolute value of the county-pair difference in the unemployment rate; in percentage points; source: US Bureau of Labor Statistics.
Gross migration	Sum of the number of people migrating from source to destination county and vice versa; source: US Census Bureau.
Gross trade	Gross value of trade between source and destination county; in million USD; source: US Census Bureau.
% commuting	Share of source county population which commutes to the destination county; in percent; source: US Census Bureau.
Same state	Binary variable equal one if source and destination county are located in the same state; source: NBER's county distance database.
Common border	Binary variable equal one if source and destination county are direct neighbors; source: US Census Bureau.

(*table continued on next page*)

Table A1 - *continued*

Variable name	Description
Further variables	
Large banks	Binary variable equal one if banks' average total assets per USD of loans provided from source to destination county is larger than the banks' median total assets per USD of cross-county loans; source of balance sheet data: FDIC.
Population control	One over the product of the source and destination county populations; source: US Census Bureau.
Physical distance (pop.-weighted centroids)	Great circle distance in miles based on *population-weighted* county centroids; source: US Census Bureau, own calculation.
Highway travel cost	Costs of traveling from source to destination county via highways; source: Oak Ridge National Laboraty's National Transportation Center.
All modes travel cost	Costs of traveling from source to destination county via the cheapest combination of highways, roads, and waterways; source: Oak Ridge National Laboraty's National Transportation Center.
Culture: ethnic ancestry	Index quantifying a county pair's cultural distance based on ethnic ancestries; calculation described in Section 2; source: own calculation.
Culture: racial origin	Index quantifying a county pair's cultural distance based on racial origins; calculation described in Section 2; source: own calculation.
Culture: religious beliefs	Index quantifying a county pair's cultural distance based on religious beliefs; calculation described in Section 2; source: own calculation.
Culture: social environment	Index quantifying a county pair's cultural distance based on social environments; calculation described in Section 2; source: own calculation.
Vote-share differential	Absolute difference of the vote share for the republican candidate during the 2016 presidential election; source: MIT Election Data and Science Lab.

Table A2: Variables, categories, and sub-categories for the measurement of cultural distance

The table lists all variables used for the measurement of the cultural distance between counties and associates them with the four dimensions defining culture in the theoretical model in Lieske (1993): ethnic ancestry, racial origin, religious beliefs, and the social environment. This environment is further divided into sub-categories for weighting purposes. Section 2 provides a detailed explanation of the construction of our cultural distance measure.

Cultural distance				
Ethnic ancestry	**Racial origin**	**Religious beliefs**	**Social environment**	
% American	% Asian	% Black Protestant	**Age**	**Mobility**
% British	% black	% Evangelical Protestant	% 19 or younger	% 5 years not moved
% Eastern European	% Hispanic	% Mainline Protestant	% 20 to 29	**Occupation**
% French	% Native American	% Catholic	% 30 to 64	% agriculture
% German	% white	% Mormon	% over 64	% construction
% Greek		% Orthodox	**Education**	% manufacturing
% Irish			% $\geq$ college degree	% service
% Italian			% $<$ high-school diploma	**Population**
% Northern European			**Family**	% urban
% Russian			% two-parent families	% total
% Subsaharan African			% females in labor force	**Racial diversity**
			Income inequality	Gini coefficient of racial origins
			Gini coefficient	

Table A3: Descriptive statistics at county-pair level

The table displays statistics for the mortgage loan sample used for the distinction between supply and demand-side effects (see Table 4, columns 2 to 4). Table 1 provides descriptive statistics for the sample used in all other analyses. All variables are at the county-pair level. For instance, "volume of SME loans" captures the volume of loans provided by all bank branches in a source county to all small and medium-sized enterprises in the destination county. Variable definitions and data sources are summarized in Table A1.

	N	Mean	Median	SD	Min	Max
Volume of mortgage loans [thousand USD]	34,483	1,559	242	8,667	0	412,072
log(Volume of mortgage loans)	34,483	10.5	12.4	5.2	0.0	19.8
Loan demand [thousand USD]	34,483	2,206	360	11,391	0	540,533
log(Loan demand)	34,483	13.0	12.8	1.5	0.0	20.1
Social connectedness	34,483	33	4	42	0	100
Physical distance [miles]	34,483	452	272	501	5	4,898
log(Physcial distance)	34,483	5.4	5.6	1.4	1.6	8.5
Cultural distance	34,483	14	13	7	0	47
GDP growth differential	34,483	3.2	2.4	3.2	0.0	37.1
Unemployment differential	34,483	1.3	1.0	1.2	0.0	20.3
Gross migration	34,483	177	1	451	0	2,327
% commuting	34,483	0.4	0.0	1.7	0.0	33.0
Gross trade [million USD]	34,483	68	21	111	0	1,056
Same state	34,483	0.3	0.0	0.5	0.0	1.0
Common border	34,483	0.2	0.0	0.4	0.0	1.0

Table A4: The role of distance and connectedness: alternative clustering

The table compares estimates for our baseline regression using alternative clustering approaches. Column 1 restates our baseline regression reported in column 4 of Table 2, where standard errors are clustered at the source and destination county levels. In column 2, standard errors are clustered at the source and destination state levels. Column 3 accounts for the dyadic structure of our data by following the clustering approach in Cameron and Miller (2014). Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)
Dependent variable:	log(Volume of SME loans)		
Clustering:	Source and destination county	Source and destination state	Dyadic
Social connectedness	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Physical distance	-0.262***	-0.262***	-0.262***
	(0.048)	(0.046)	(0.048)
Cultural distance	-0.015**	-0.015**	-0.015**
	(0.007)	(0.006)	(0.007)
Source county FE	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684
Adj. R^2	0.524	0.524	0.524
Adj. R^2 within	0.131	0.131	0.131

Table A5: Instrumental variable approaches: first-stage regressions

This table reports the first stage regressions of the instrumental variable estimations reported in Table 3. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)	(4)
Dependent variable:	Social connectedness			
Same highway	2.951***			
	(0.504)			
Years since highway construction		0.055***		
		(0.010)		
Historical travel costs			4.070***	
			(0.348)	
Relative Facebook county rank				161.506***
				(12.629)
Physical distance	-10.914***	-10.919***	-23.849***	-16.840***
	(1.137)	(1.137)	(1.196)	(1.074)
Cultural distance	-0.418***	-0.417***	-0.361***	-0.847***
	(0.052)	(0.052)	(0.055)	(0.073)
GDP growth differential	0.010	0.011	-0.054	-0.012
	(0.065)	(0.065)	(0.048)	(0.086)
Unemployment differential	-0.772***	-0.772***	-0.689***	-0.685***
	(0.209)	(0.209)	(0.169)	(0.205)
Gross migration	0.005***	0.005***	0.005***	-0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
Gross trade	-0.003	-0.003	-0.002	0.010***
	(0.003)	(0.003)	(0.003)	(0.004)
% commuting	0.193	0.206	-0.177	-0.232
	(0.141)	(0.142)	(0.128)	(0.258)
Same state	27.002***	27.009***	23.308***	30.760***
	(1.820)	(1.820)	(1.460)	(1.671)
Common border	20.023***	20.030***	9.763***	20.338***
	(1.367)	(1.367)	(1.372)	(1.205)
Population control				-0.000***
				(0.000)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	56,223	56,852
Adj. R^2	0.858	0.858	0.881	0.828
Adj. R^2 within	0.676	0.676	0.733	0.569

Table A6: Instrumental variable approaches: alternative specifications

Columns 1 and 2 re-estimate the instrumental variable regressions reported in Table 3, columns 1 and 2 while excluding counties in wich the share of the population living in rural areas is above the 75^{th} percentile of this rural population share. In column 3, we repeat our instrumental variable regression based on the initial facebook rollout, but use *relative* social connectedness instead of the absolute measure in Table 3, column 4. Control variables are the GDP growth and unemployment differentials, gross migration and trade, as well as same state and common border indicator variables. First-stage coefficients, p-values, and F-statistics of the instruments are reported in the bottom section of the table. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)	(3)
Dependent variable:	log(volume of SME loans)		
Instrument:	same highway	years since construction	facebook rollout
Social connectedness	0.029*	0.031*	0.122*
	(0.015)	(0.016)	(0.074)
Physical distance	-0.018	0.002	0.780
	(0.173)	(0.183)	(0.655)
Cultural distance	-0.003	-0.002	0.043
	(0.009)	(0.010)	(0.033)
Source county FE	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
No. of obs.	49,725	49,725	56,852
Adj. R^2	0.497	0.493	0.041
Adj. R^2 within	0.103	0.096	-0.845
Instrument (1st stage)	2.88***	0.054***	16.85***
	(0.000)	(0.000)	(0.001)
F-value (1st stage)	28.996	26.028	10.3

Table A7: The role of distance and connectedness: SME loans vs. mortgage loans

Columns 1 and 2 re-estimate the instrumental variable regressions reported in Table 3, columns 1 and 2 while excluding counties in wich the share of the population living in rural areas is above the 75^{th} percentile of this rural population share. In column 3, we repeat our instrumental variable regression based on the initial facebook rollout, but use *relative* social connectedness instead of the absolute measure in Table 3, column 4. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. First-stage coefficients, p-values, and F-statistics of the instruments are reported in the bottom section of the table. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

Dep. var.: log(Volume of ...)	(1) SME loans	(2) Mortgage loans
Social connectedness	0.007***	0.009***
	(0.001)	(0.003)
Physical distance	-0.263***	-0.486***
	(0.048)	(0.109)
Cultural distance	-0.014*	-0.035**
	(0.007)	(0.015)
GDP growth differential	-0.006	-0.016
	(0.008)	(0.022)
Unemployment differential	-0.004	-0.044
	(0.020)	(0.049)
Gross migration	0.000***	0.000
	(0.000)	(0.000)
Gross trade	0.000	-0.001
	(0.000)	(0.001)
% commuting	-0.064***	-0.077***
	(0.010)	(0.018)
Same state	0.092	0.426**
	(0.068)	(0.183)
Common border	0.821***	1.747***
	(0.063)	(0.150)
Source county FE	Yes	Yes
Destination county FE	Yes	Yes
No. of obs.	66,684	34,483
Adj. R^2	0.525	0.157
Adj. R^2 within	0.132	0.052
Standardized beta coefficients		
Social connectedness	0.12	0.07
Physical distance	-0.17	-0.13
Cultural distance	-0.05	-0.05

Table A8: The role of distance and connectedness: headquarter location

Column 1 restates the baseline regression from Table 2, column 4, where bank locations are based on their branch network. In column 2, we re-estimate our baseline regression, but determine bank locations based on their headquarters. Control variables are the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Variable definitions are provided in Table A1. The parentheses contain standard errors clustered at the source county and destination county levels. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	(1)	(2)
Dependent variable:	log(Volume of SME loans)	
Bank location:	Branch location	Headquarter location
Social connectedness	0.007***	0.012***
	(0.000)	(0.000)
Physical distance	-0.263***	-0.370***
	(0.000)	(0.000)
Cultural distance	-0.014*	-0.026***
	(0.056)	(0.004)
Source county FE	Yes	Yes
Destination county FE	Yes	Yes
Controls	Yes	Yes
No. of obs.	66,684	73,305
Adj. R^2	0.525	0.545
Adj. R^2 within	0.132	0.185
Standardized beta coefficients		
Social connectedness	0.15	0.12
Physical distance	-0.18	-0.17
Cultural distance	-0.09	-0.05