ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 033

The Role of Social Networks in Bank Lending

Oliver Rehbein Simon Rother

October 2020 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged.

The Role of Social Networks in Bank Lending*

Oliver Rehbein†
University of Bonn

Simon Rother‡
University of Bonn

October 1, 2020
(Please click here for the latest version)

Abstract

This paper analyzes social connectedness as an information channel in bank lending. We move beyond the inefficient lending between peers in exclusive networks by exploiting Facebook data that reflect social ties within the U.S. population. After accounting for physical and cultural distances, social connectedness increases cross-county lending, especially when lending requires more information and screening incentives are intact. On average, a standard-deviation increase in social connectedness increases cross-county lending by 24.5%, which offsets the lending barrier posed by 600 miles between borrower and lender. While the ex-ante risk of a loan is unrelated to social connectedness, borrowers from well-connected counties cause smaller losses if they default. Borrowers' counties tend to profit from their social proximity to bank lending, as GDP growth and employment increase with social proximity. Our results reveal the important role of social connectedness in bank lending, partly explain the large effects of physical distance, and suggest implications for antitrust policies.

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

*We thank René Bernard (discussant), Martin Brown, Hans Degryse, Ralph De Haas, Florian Heider, Mikael Homanen (discussant), Luc Laeven, Yueran Ma, Ralf Meisenzahl, Alexandra Niessen-Ruenzi, Steven Ongena, Loriana Pelizzon, Farzad Saidi, Martin Schmalz, Isabel Schnabel, Sascha Steffen, Johannes Stroebel, and seminar participants in Bonn and Zurich for valuable comments and suggestions. We also thank participants of the annual meeting of the German Finance Association (DGF) in Essen, the SAFE workshop on Household Finance in Frankfurt, the conference of the International Banking, Economics, and Finance Association (IBEFA) in San Diego, and the workshop on Banking in Mannheim for valuable feedback. Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Projects C03 and C04) and through EXC 2126/1-390838866 under Germany's Excellence Strategy is gratefully acknowledged. A previous version of this paper circulated under the title "Distance in Bank Lending: The Role of Social Networks."

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

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

1 Introduction

Serving as an information channel, real-world social networks can help to overcome information frictions and, where they do, improve economic outcomes. In bank lending, the information frictions between borrower and lender are particularly important. They are costly to overcome and provide a key justification for the very existence of banks (Diamond, 1984; Boot, 2000). As soft information enters banks' lending decisions, strong social ties appear likely to result in a more efficient credit intermediation process by reducing the need for and the cost of information acquisition about borrowers or their local economic environments. Yet, in the context of bank lending, social networks are predominantly associated with negative consequences such as inefficient loan allocations and impaired loan performance (Khwaja and Mian, 2005; Haselmann, Schoenherr, and Vig, 2018). These negative consequences result from crony lending between peers in exclusive networks. While it is important to be aware of this dark side of social ties, it remains unclear whether social networks, when defined more broadly, can facilitate banks' access to information and, thereby, improve bank lending. This question is of particular interest as social networks become increasingly widespread and people exchange information ever more rapidly.

We exploit a unique dataset that reflects social ties within the U.S. population. Based on this dataset, we analyze the role of social connectedness as an information channel in bank lending. Specifically, we ask three questions. First, how does social connectedness affect the allocation of loans? Second, is this effect associated with an information channel? And third, what are the consequences of these lending decisions for borrowers and banks? Our results suggest that social connectedness increases lending in a way that is in line with an information channel which benefits borrowers and banks. To account for prominent factors which aggravate information frictions, we control for the physical and cultural distances between borrower and lender throughout the analyses.

As such, this paper also offers new insights into the role of physical distance (Degryse and Ongena, 2005; Agarwal and Hauswald, 2010) and cultural differences (Giannetti and Yafeh, 2012; Fisman, Paravisini, and Vig, 2017) in bank lending. We make use of the social connectedness channel to analyze to what extent the effect of physical distance can be attributed to soft information rather than to transportation costs. Information flows through social networks also offer a rationale for banks' limited ability to collect soft information at large distances, as the density of networks decreases with distance. Additionally, we study whether social connectedness can compensate for the lending barriers posed by physical and cultural distance and analyze the interactions of the effects of connectedness and distances.

To measure social connectedness, we leverage a recent dataset on Facebook friendship links in the United States (Bailey, Cao, Kuchler, and Stroebel, 2018), where the use of Facebook is pervasive. As of 2019, the share of monthly active users amounts to 75% of the total U.S. population. Facebook friendship links mostly correspond to real-world networks of relatives, colleagues, business partners, and friends. In 2020, COVID-19 infections spread along the social ties reflected by the data (Kuchler, Russel, and Stroebel, 2020). Hence, the data allow a comprehensive assessment of real-world social connections in which information can be exchanged both online and in person. The data are aggregated at the county-pair level and provide the relative probability of a person in county A being acquainted with a person in county B.

We supplement this information with data on loans to small and medium-sized enterprises (SMEs) from the Community Reinvestment Act (CRA) as well as mortgage lending from the Home Mortgage Disclosure Act (HMDA). Small firms tend to be more opaque borrowers for whom soft information is more important during the credit intermediation process. The mortgage-loan data also allow us to analyze the riskiness and the performance of loans. Our measure of cultural distance builds on the theoretical models of regional subcultures in Elazar (1984) and Lieske (1993). These models define culture as a combination of a person's ethnic ancestry, religious beliefs, racial origin, and the structure of their social environment. We collect a wide array of variables on these categories to compute the cultural distance between U.S. counties. The resulting measure is theory based, considers several dimensions of cultural identity, and corresponds to well-known patterns.

Our results reveal that social connectedness significantly increases county-to-county lending. In our baseline regression, a standard-deviation increase in social connectedness is associated with a 24.5% increase in SME loan volumes. For mortgage lending to households, we find a weaker effect, which is in line with an information channel, as the credit intermediation process is more standardized for mortgage loans and SMEs are more opaque borrowers. Social connectedness also increases the probability of bank lending from a source to a destination county. The effect of social connectedness is distinct from physical and cultural distances, for which we control in all regressions.

Interestingly, social connectedness explains part of the effect of physical distance. In line with the literature, loan volumes decrease with physical distance. However, accounting for social connectedness significantly shrinks this effect. The information that runs through social networks thus provides an explanation for the large distance effects in the literature. In economic terms, the opposing effects of social connectedness and physical distance are similarly large. A standard-deviation increase in social connectedness compensates for more than 600 miles of additional distance between borrower and lender. Hence, social connectedness can significantly help to overcome the lending barriers posed by physical distance. Similar to physical distance, cultural distance is also associated with lower loan volumes. However, this negative effect of cultural differences entirely disappears in the presence of sufficiently close social ties.

The relevance of social connectedness increases with banks' need for information. Small banks, which have a less standardized credit intermediation process, experience a larger increase of loan volumes in social connectedness. Similarly, social connectedness is more important when borrowers' creditworthiness is more challenging to evaluate because of a higher exposure to industry volatility or their local economic environments exhibiting a strong boom or bust. The relevance of social connectedness also increases if the local economic development in a borrower's region is very different from that in the bank's region, which constitutes an additional case of high information need. Conversely, social connectedness increases cross-county lending less for loans with reduced screening incentives due to government guarantees or securitization. These findings strongly suggest that social connectedness affects loan allocations because of an information channel.

Based on a loan-level analysis, we find no evidence of social connectedness being associated with more risky lending. Borrower credit scores and loan-to-value ratios are not significantly related to social connectedness. Borrowers from well-connected counties pay lower interest rates, which is in line with a lower cost of information acquisition for banks. While delinquency rates do not significantly vary with social connectedness, actual defaults lead to lower losses: controlling for the initial loan amount and further characteristics, a standard-deviation increase in social connectedness reduces the outstanding amount of defaulting loans by 80%. Banks thus profit from better performance of loans to borrowers from well-connected regions, which further supports the notion that social connectedness facilitates banks' access to information.

From the borrowers' perspective, social proximity to bank capital is highly valuable. In addition to more lending, counties with higher social proximity to bank capital experience higher GDP growth and more employment. Specifically, one standard deviation higher social connectedness is associated with 0.85 percentage points higher GDP growth and 0.5% higher employment. Regions with a high percentage of small firms, which rely more on bank loans for financing, profit particularly strongly from their social proximity to banks. The analysis of real effects thus provides no evidence of a connectedness-driven financing of negative net-present-value (NPV) projects.

Our baseline results on loan allocations 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 county-to-county lending, including migration, commuting behavior, and trade. To improve identification, we construct a panel dataset to account for source-county-time and destination-county-time fixed effects, and estimate instrumental-variable regressions that exploit historical travel costs and the quasi-random rollout of Facebook as instruments. The estimates further corroborate our results on social connectedness. Moreover, our results are robust to alternative measures of physical and cultural distance, which emphasizes that the role of social connectedness in bank lending is distinct from these distances. The results cannot be explained by differences in state-level regulation either, as they apply within and across states. Lastly, the results are robust to alternative approaches to the clustering of standard errors, including dyadic clustering.

Overall, social connectedness increases bank lending, especially when banks have a high need for information and screening incentives are intact. Banks and especially borrowers profit from the resulting loan allocations. Hence, social networks, when defined broadly, can help to overcome information frictions and improve bank lending. These findings suggest three implications. First, regulators may want to take social connections into account in antitrust decisions. Whereas distance remains an important factor, a high concentration of lenders in a geographical area appears less problematic if it has close social ties to regions in which other banks are located. Second, social connectedness may help to explain the trend toward geographically more dispersed banks over the past decade. Social networks drive loan allocations and the networks have become increasingly widespread with an ever more rapid exchange of information. Third, banks may expand into a region more efficiently when strategically employing well-connected loan officers. While the literature shows that direct bonds between the borrower and lender lead to inefficient lending decisions, social ties to a borrower's region facilitate a bank's access to information and, on average, result in more efficient loan allocations.

Our paper is embedded in a broad literature on the importance of social networks for economic outcomes.[1] Jackson (2011) provides a comprehensive overview. Social networks are known to affect the quality of information flows and trust (Granovetter, 2005), thereby shaping economic outcomes. Several studies analyze how investment behavior depends on social connections.[2] Yet, despite banks actively relying on soft information for their lending decisions (Uchida, Udell, and Yamori, 2012; Liberti, 2018; Gropp and Güttler, 2018), and despite particularly pronounced information frictions between borrowers and lenders, the relevance of social connections as an information channel for bank-lending decisions has hardly been analyzed. La Porta, López de Silanes, and Shleifer (2002) and Khwaja and Mian (2005) show that political connections drive lending decisions. Haselmann, Schoenherr, and Vig (2018) show that bank directors extend more inefficient credit to members of their elite social club.[3] We contribute to this literature by analyzing the ubiquitous social network

[1]For instance, Duflo and Saez (2003) analyze the role of social networks in individual retirement decisions. Also see Ioannides and Datcher-Loury's (2004) discussion of the role of social networks in labor markets. Nguyen (2012) and Kramarz and Thesmar (2013) look at social networks within boards and in the upper management of firms. Chaney (2014) and Bailey, Farrell, Kuchler, and Stroebel (2019) investigate the role of networks in international trade. Bailey, Cao, Kuchler, Stroebel, and Wong (2018) demonstrate that information about house price developments spreads along socially connected individuals. Bailey, Johnston, Kuchler, Stroebel, and Wong (2019) study the role of social networks for the adoption of new products.

[2]See Kelly and Ó Gráda (2000), Hong, Kubik, and Stein (2004), Hong, Kubik, and Stein (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 social ties, which helps them to outperform the market. Kuchler, Li, Peng, Stroebel, and Zhou (2020) show that institutional investors invest more in firms located in areas to which they are well connected, but these investors do not achieve superior returns.

[3]Lin, Prabhala, and Viswanathan (2013) analyze data from a peer-to-peer lending platform and show that lenders' decisions depend on the behavior of a borrower's online friends.

that spans a society rather than an elite club. Through this broad network, loan officers can receive information about a borrower or their local economic environment without having a direct personal connection to that borrower and, hence, without necessarily receiving a private benefit from crony lending. In line with this difference in the nature of the network, we find that social connectedness increases lending because of an information channel and in a way that is beneficial for banks and the real economy.

Furthermore, our paper relates to the literature on relationship banking (see, for example, Boot, 2000; Kysucky and Norden, 2015) in general and on the effects of physical distance in particular. The effect of physical distance on lending outcomes is highlighted by a long list of influential studies.[4] While transportation costs are one potential explanation for the relevance of physical distance, parts of the literature explicitly rationalize the physical distance effects with banks being only able to collect soft information locally (see, for example, Agarwal and Hauswald, 2010).[5] Given the recent advances in information technology, the collection of soft information may be hindered more by differences in social and cultural backgrounds than physical transportation costs. We contribute to this literature by showing that the information flowing along social ties partly explains the large effects of physical distance and that social connectedness can compensate for the lending barrier posed by distance. These findings also speak to competition policies. Markets are often defined geographically, such that physical distance is a main driver of competition (Degryse and Ongena, 2005; Granja, Leuz, and Rajan, 2018), but our results illustrate that sociocultural factors also determine loan allocations.

Lastly, our paper connects to studies of cultural differences between borrowers and lenders. Beck, Degryse, de Haas, and van Horen (2018) highlight that foreign banks have disadvantages in collecting local information, which may be due to cultural differences. From a firm's perspective, such disadvantages can be reduced by owning more foreign assets (Houston, Itzkowitz, and Naranjo, 2017). Giannetti and Yafeh (2012) demonstrate that cultural differences between countries affect

[4]The non-exhaustive list includes Petersen and Rajan (2002), Berger, Miller, Petersen, Rajan, and Stein (2005), Degryse and Ongena (2005), Brevoort and Hannan (2006), Mian (2006), DeYoung, Glennon, and Nigro (2008), Agarwal and Hauswald (2010), Hollander and Verriest (2016), Beck, Ongena, and Şendeniz-Yüncü (2019), Nguyen (2019).

[5]Also highlighting the relevance of information for lending distances, Degryse, Laeven, and Ongena (2009) find that banks with inferior information technology lend at shorter distances.

cross-country lending. Based on data from India, Fisman, Paravisini, and Vig (2017) show that more loans are extended and repayment rates increase if the loan officer and the lender are similar in terms of caste and religion, which suggests that cultural differences aggravate information frictions in bank lending. Our findings are in line with these studies in that cultural distance constitutes a lending barrier. We contribute to this literature in two regards. First, we introduce a new measure of cultural differences between counties in the United States. This measure is theory-based, considers several dimensions of cultural identity, and corresponds to well-known patterns. Second, we analyze the interplay between social connectedness and cultural differences: cultural distance constitutes a lending barrier even when controlling for social connectedness, but the negative effects of cultural distance disappear in the case of sufficiently close social ties.

2 Empirical strategy

We conduct our analysis in three steps. First, we analyze how social connectedness affects the allocation of bank lending and how this effect depends on the information sensitivity of loans. Second, we further explore the information channel and assess consequences of the altered loan allocations based on a loan-level analysis of the ex-ante lending risk and the ex-post loan performance. Third, we analyze consequences for borrowers by studying the real effects of counties' social proximity to bank capital. Subsequently, we describe our empirical strategy in detail. The data are described in Section 3.

2.1 Allocation of bank lending

Baseline specification Our main variable of interest measures the strength of social connections between U.S. counties. In our baseline regressions, we explain bank lending from branches in source county i to borrowers in destination county j by the counties' social connectedness while controlling for their physical distance (in logs), their cultural distance, further county-pair-specific control variables, and source and destination county fixed effects.

$$\begin{aligned} \text{bank lending}_{i,j} = {} & \beta_1 \cdot \text{social connectedness}_{i,j} \\ & + \gamma_1 \cdot \ln(\text{physical distance})_{i,j} + \gamma_2 \cdot \text{cultural distance}_{i,j} \\ & + \gamma_3 \cdot \text{county-pair-level controls}_{i,j} + \alpha_i + \alpha_j + \epsilon_{i,j} \end{aligned} \tag{1}$$

The dependent variable is the volume of loans (in logs). In additional regressions, we analyze the probability of a lending relationship. The county-pair-specific control variables account for the GDP growth and unemployment differentials (in absolute terms), gross trade and migration, the share of the commuting population, and same-state and common-border indicator variables. We include the unemployment rate and GDP growth differentials, because banks may take into account how different economic conditions are compared to their home market. The trade volumes account for the interconnectedness of industries. Migration and commuting may simultaneously affect bank lending and social connectedness. The same-state indicator accounts for regulation that may hinder banks in expanding their business across state borders. Standard errors are clustered at the source- and destination-county levels.[6] Even though it appears unlikely that a significant share of the social connections in the population emerges due to bank lending, we lag all explanatory variables by one year to mitigate reverse causality concerns.

Additional identification Since counties' connectedness and distances are time-invariant or at least highly persistent, our baseline regression is based on cross-sectional data. The results do, however, also hold up in a longer panel, where we include source-county-time and destination-county-time fixed effects to control, for instance, for the time-varying economic conditions in the source county and credit demand in the destination county. To provide additional identification for our cross-sectional baseline setting, we introduce several instrumental variable approaches in Section 4.1 based on historical travel costs and the quasi-random rollout of Facebook.

Information sensitivity To explore whether the effect of social connectedness is related to information, we analyze heterogeneities across the information sensitivity of loans. To this end, we interact social connectedness in our baseline specification (Equation 1) with measures of bank types, borrower types, the borrowers' local economic environments, and loan types. This allows us to assess how the effect of social connectedness depends on banks' need for information and their screening incentives.

[6]The results are robust to state-level and dyadic clustering (see Table OA1 in the Online Appendix).

2.2 Lending risk and loan performance

To further distinguish between crony lending and the information channel and to learn about the consequences of social connectedness affecting banks' lending decisions, we analyze the riskiness of loans from an ex-ante and an ex-post perspective based on a loan-level sample. To this end, we estimate loan-level regressions that explain the riskiness of loan l originated in year t by bank b by the social connectedness between source county i (=branch location) and destination county j (=borrower location) while controlling for physical and cultural distance, additional loan characteristics, and bank and origination-year fixed effects.

$$\begin{aligned}\text{riskiness}_l = {} & \beta_1 \cdot \text{social connectedness}_{i,j} \\ & + \gamma_1 \cdot \ln(\text{physical distance})_{i,j} + \gamma_2 \cdot \text{cultural distance}_{i,j} \\ & + \gamma_3 \cdot \text{additional loan characteristics}_l + \alpha_b + \alpha_t + \epsilon_l \end{aligned} \tag{2}$$

Our measures of ex-ante riskiness are the borrower's credit score and the loan-to-value ratio. The ex-post loan performance is based on delinquency rates and the remaining outstanding amount of defaulting loans. We also analyze the relationship between social connectedness and the loans' interest rates. The additional loan characteristics control for the original loan amount (in logs), the debt-to-income ratio, and whether the borrower is a first-time home buyer.

2.3 Real effects

To assess implications of social connectedness from a borrower perspective, we estimate the real effects of borrower counties' social proximity to bank capital. Specifically, we regress an outcome in county c on the county's social proximity while controlling for its physical and cultural proximity, county- and state-time fixed effects, and additional control variables.

$$\begin{aligned}\text{outcome}_{c,t} = {} & \beta_1 \cdot \text{social proximity}_{c,t-1} \\ & + \gamma_1 \cdot \text{physical proximity}_{c,t-1} + \gamma_2 \cdot \text{cultural proximity}_{c,t-1} \\ & + \gamma_3 \cdot \text{additional control variables}_{c,t-1} + \alpha_c + \alpha_{s,t} + \epsilon_{c,t} \end{aligned} \tag{3}$$

The outcome variables are loan volumes (in logs), real GDP growth, and employment (in logs). The additional control variables account for industry shares, commuting, and migration. All explanatory variables enter the regressions lagged by one year.

3 Data

For our analyses, we construct three main datasets. First, to study loan allocations, we collect data on county-to-county lending, which corresponds to the level of observation of our main explanatory variable, social connectedness. Second, we build a loan-level sample to analyze the riskiness and performance of loans. Third, we construct a county-level sample to study real effects. Subsequently, we discuss each dataset in detail. Table A1 in the appendix summarizes the data sources and provides variable definitions.

3.1 Allocation of bank lending

Social connectedness Our measure of social connectedness is based on a 2016 cross section of the universe of Facebook friendship links from the United States, introduced by Bailey, Cao, Kuchler, and Stroebel (2018). For confidentiality, the dataset is aggregated at the county-pair level and scaled by an unknown factor. It is normalized by the product of the county-pair population. We winsorize the variable at the 99^{th} percentile to account for outliers in the distribution and rescale it to range between 0 and 100, where 100 is the maximum social connectedness between counties. This allows us to interpret the final variable as the relative probability that a person in county A is acquainted with a person in county B, measured in percent of the maximum probability between county pairs. The measure reveals the structure of real-world social networks as the use of Facebook is pervasive across the U.S. population and Facebook friendship links predominantly correspond to real-world connections between relatives, friends, colleagues, and business partners (Bailey, Cao, Kuchler, and Stroebel, 2018), as was also illustrated by COVID-19 infections spreading along the social ties within the data (Kuchler, Russel, and Stroebel, 2020).

Figure 1 illustrates the variation in social connectedness based on the example of Montgomery County, OH, which is representative in our data in terms of its correlation with physical and cultural distances. The largest city in this county is Dayton. Areas colored in dark blue exhibit the highest social connectedness, light colors represent low connectedness. Many counties in or near Ohio are socially well-connected to Montgomery County. However, there are also significant connections to more distant areas such as Southern Florida, parts of Colorado and the East Coast, and a number of individual counties scattered across the United States. The high connectedness to Southern Florida is in line with its status as a prime destination for retirement and tourism among people in the northeast of the United States. The close social ties to the various more scattered counties also correlate with a common factor: the largest employer in Montgomery County is the Wright-Patterson Air Force base and most of these closely connected counties also host Air Force bases. For example, the lone dark-blue spot in Idaho is Elmore County, which hosts the Mountain Home Air Force Base that accounts for 15% of the county's population.[7] While social connectedness and physical distance are significantly correlated (-0.49), the figure illustrates that counties within the same area can differ strongly in their social connectedness, which is partly determined by highly idiosyncratic county characteristics and allows us to estimate the effect of social connectedness while controlling for distance.

– Figure 1 around here –

Lending For our main analyses, we obtain data on lending to small and medium-sized enterprises collected under the Community Reinvestment Act (CRA). These firms are opaque borrowers, making soft information particularly important for banks. The dataset exhibits a broad coverage and comprises *newly originated* loans which amount to over 230 billion USD for 2017.[8] Additional regressions rely on mortgage-lending data collected under the Home Mortgage Disclosure Act (HMDA), which also indicates whether a loan is backed by government guarantees or sold for securitization. The mortgage-lending data comprise 14.3 million loans originated from 5,852 financial institutions in 2017.[9] Both datasets are available through the Federal Financial Institutions

[7] The non-exhaustive list of additional examples include the Minot Air Force Base in Ward County, North Dakota; the US Air Force Academy in El Paso County, Colorado; the Ellsworth Air Force Base in Pennington County, South Dakota; the Altus Air Force Base in Jackson County, Oklahoma; and the Creech and Nellis Air Force Bases in Clark County, Nevada.

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

[9] 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`.

Examination Council (FFIEC). We assign each loan to the lending bank's branch located closest to the borrower based on branch locations provided by the FDIC.[10] Using the borrower locations reported in the datasets, we aggregate information on total loan volumes at the county-pair level.

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 U.S. Census Bureau and usually correspond to a county's geographical center. In robustness checks, we consider the great-circle distance between population-weighted county centroids (U.S. Census Bureau) and traveling costs by highway, rail, and waterways (National Transportation Center).

Cultural distance We construct a measure of cultural distance at the U.S. 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 their social environment. To operationalize the models, we collect 39 variables in these four categories from the 2010 U.S. Census, the 2010 American Community Survey, and the 2010 U.S. Religious Congregations and Membership Study. Table A2 in the appendix provides an overview of the variables in each category and, in the case of the social environment, subcategories. Figure 2 illustrates the variation in the data based on a cluster analysis of the principal components of the 39 culture variables. The resulting pattern corresponds to well-known historical patterns such as the so-called Black Belt (dark green area) in the southeast of the United States.

– Figure 2 around here –

To measure the cultural distance between counties, we calculate the absolute difference per county pair for each variable and sum these differences across all variables of one subcategory.

[10]For robustness, we alternatively determine source locations based on banks' headquarters. The findings are robust with slightly smaller estimates in economic terms (see Table OA2 in the Online Appendix), which is plausible as bank headquarters can always obtain information from their branches.

Afterward, we sum across subcategories and, finally, across categories. To ensure equal contribution to the variation within the final variable, within categories, and within subcategories, we normalize every summand to mean zero and variance one before calculating the sum. As we analyze in Section 5, our results are robust to including all 39 culture variables individually, but the aggregation allows for a meaningful interpretation of cultural differences.[11] We scale the final variable to range between 0 and 100 so that it can be interpreted as the cultural distance as a percentage of the maximum cultural distance between any two U.S. counties. As expected, cultural distance correlates negatively to social connectedness (-0.17) and positively to physical distance (0.38). When regressing social connectedness on physical and cultural distance, the distances explain 24% of the variation in social connectedness, such that there is sufficient remaining variation to analyze the effects of social connectedness while controlling for physical and cultural distance.

Further main covariates We collect gross-commuting and gross-migration population shares at the county-pair level (U.S. 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 (U.S. 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 (Bureau of Economic Analysis) and unemployment rates (Bureau of Labor Statistics) and calculate the absolute value of the county-to-county difference for each of the two variables.

Final dataset Most of the over 9.5 million county pairs in the United States exhibit no cross-county lending. To avoid that the dependent variable mostly equals zero, we restrict our sample to county pairs with at least one cross-county loan. The lower number of observations also leaves us on the conservative side with respect to the statistical significance of our estimates. Our results hold in a dataset that includes all county pairs, which we also use to analyze the probability of lending between county pairs.

[11] We also exploit differences in voting patterns as a proxy for cultural differences, and account for the cultural heterogeneity within counties (see Section 5).

Our final dataset comprises lending to SMEs from 1,944 source counties in 50 states to 3,086 destination counties in 50 states, resulting in a total of 66,684 county pairs. Mortgage lending takes place between 34,483 county pairs, but only 8,532 county pairs simultaneously exhibit both types of lending. Subsequently, we discuss summary statistics for our main sample of SME lending (Table 1), but the variation in the mortgage-lending sample is similar (compare Table OA3 in the Online Appendix). The median volume of cross-county lending is close to 140,000 USD and the distribution is highly skewed as loan volumes can amount to almost 1.3 billion USD. The median social connectedness is only 2% of the maximum social connectedness between counties and social connectedness varies greatly, as the standard deviation equals 35. The median county-to-county distance is slightly above 400 miles. The median cultural distance equals 16% of the maximum cultural distance between counties. The GDP growth and unemployment differentials, gross trade, gross migration, and gross commuting show a large range, reflecting the variety of economic and structural conditions across regions in the United States.

– Table 1 around here –

3.2 Lending risk and loan performance

Sample construction and additional data sources We merge the HMDA data with Fannie Mae's and Freddie Mac's Single Family Loan-Level Datasets, which contain detailed information on borrower characteristics, loan characteristics, and loan performance of 30-year fixed rate mortgages acquired by these two institutions. With the exception of the loan performance measures, all variables used in our analyses are as of the time of origination. The Single Family data and the HMDA data do not contain a unique identifier. We follow the strategy in Saadi (2020) to merge only uniquely identified sets of loans based on observable characteristics.[12] This restricts the sample to a representative subset of the Single Family datasets.

Final dataset and main additional variables Our final sample contains 20,760 loans originated between 2000 and 2008. We observe the performance of these loans until the end of 2018.

[12]Specifically, origination year, original loan amount, loan purpose, occupancy type, and three-digit ZIP code.

Table 2 reports summary statistics. The distribution of the connectedness and distance variables, still based on the locations of the borrower and of the branch of the bank that originally originated the loan, is similar to our baseline sample of cross-county loans. Borrowers' FICO scores as of the origination date range between 300 and 835 and the median credit score is fair (638). At the median, a mortgage loan pays an interest rate of 6.4%, amounts to 150,000 USD, and finances 80% of the value of the purchased property. All three variables vary significantly. Ten percent of the loans are delinquent (90 days past due at least once), but only 0.8% of the loans have defaulted. The median outstanding amount at the time of default is 38,000 USD.

– Table 2 around here –

3.3 Real effects

Proximity to bank capital To analyze real effects, we construct a dataset at the county level. We calculate a county's proximity to banks following the centrality measure of institutional investment in Kuchler, Li, Peng, Stroebel, and Zhou (2020). Specifically, we weight social connectedness with bank capital in each county:

$$Social\ proximity\ to\ bank\ capital_{i,t} = \sum_j Social\ connectedness_{i,j} \cdot Total\ bank\ assets_{j,t}\ . \tag{4}$$

Similarly, we measure a county's physical and cultural proximities as

$$\begin{aligned} Physical\ proximity\ to\ bank\ capital_{i,t} &= \sum_j Phyiscal\ distance_{i,j}^{-1} \cdot Total\ bank\ assets_{j,t}\ . \\ Cultural\ proximity\ to\ bank\ capital_{i,t} &= \sum_j Cultural\ distance_{i,j}^{-1} \cdot Total\ bank\ assets_{j,t}\ . \end{aligned} \tag{5}$$

Total bank assets is the sum of total assets of all banks with headquarters in county j, which we obtain from the FDIC. We standardize the variables to a standard deviation of 1 to ease the interpretation of our estimates. Figure 3 illustrates counties' social proximity to bank capital as an average over the years 2009 to 2017. Dark-blue areas mark counties with high social proximity, light colors represent low proximity. The figure clearly identifies the areas around main financial hubs such as New York, Charlotte, or Minneapolis and St. Paul, but again social proximity correlates only moderately with physical (0.18) and cultural proximities (0.26). Together, the two variables explain only 9% of the variation in social proximity, allowing us to analyze the effects of social proximity while controlling for locations and culture.

– Figure 3 around here –

Additional variables and final dataset We accompany the proximity measures with our data on county-level real GDP growth, lending to SMEs, commuting, and migration. Additionally, we obtain data on employment (U.S. Bureau of Labor Statistics), industry shares (U.S. Bureau of Economic Analyses), and the percentage of small firms (U.S. Census Bureau's Statistics of U.S. Businesses). Our final sample covers 3,021 counties over the years 2009 to 2017.

Table 3 reports summary statistics. Due to the standardization, the proximity measures have a standard deviation of 1, such that their coefficients reflect the change in a dependent variable that is associated with a standard-deviation increase in a proximity measure. As expected, loan volumes, which are now aggregated at county level and include within-county lending, are larger than in the cross-county setting. In line with more volatile economic developments at more disaggregate levels, the standard deviations of GDP growth and employment are large (7.6 and 152,467). The median share of small firms is 57%.

– Table 3 around here –

4 Results

Our analysis proceeds in three steps. First, we analyze how social connectedness affects loan allocations. Second, we explore how social connectedness relates to the ex-ante lending risk and the ex-post loan performance. Third, we analyze the real effects of social proximity to banks.

4.1 Social connectedness and the allocation of bank lending

To study how social connectedness affects loan allocations, we first explain cross-county loan volumes by social connectedness and analyze how this effect relates to physical and cultural distances. Afterward, we begin to explore the information channel and assess how the effect of social connectedness depends on the information sensitivity of loans.

4.1.1 Loan allocations in light of connectedness and distances

Baseline estimates We begin our main analysis by regressing cross-county loan volumes to small and medium-sized enterprises (in logs) on social connectedness, county-pair-specific control variables, and source and destination county fixed effects. Column 1 of Table 4 reports the results. Counties with higher social connectedness exhibit more cross-county lending. The coefficient of social connectedness indicates that an increase in social connectedness by one percentage point is associated with a statistically significant increase in lending of 1.2%. The next two columns of Table 4 repeat our previous regression but include physical distance (column 2) or cultural distance (column 3) instead of social connectedness. Both variables are significantly and negatively related to bank lending. In line with the literature, physical distance (e.g., Degryse and Ongena, 2005; Agarwal and Hauswald, 2010) and cultural distance (e.g., Giannetti and Yafeh, 2012; Fisman, Paravisini, and Vig, 2017) constitute lending barriers. In contrast, social connectedness increases bank lending.

The coefficients of the control variables have the expected signs (see Table OA4 in the Online Appendix). Loan volumes are higher within states and in neighboring counties. The GDP growth and unemployment differentials are associated with lower loan volumes, although the coefficients are not statistically significant. Lastly, loan volumes tend to increase with gross trade, commuting, and migration between the counties.

Column 4 of Table 4 reports the results for our baseline regression (see Equation 1), which simultaneously includes social connectedness and physical and cultural distance together with the control variables and fixed effects. The coefficient of social connectedness decreases but remains positive and statistically significant. Hence, bank lending increases with social connectedness and this relationship is distinct from physical and cultural distance. The coefficients of physical and cultural distance also become smaller (in absolute terms; see bottom part of Table 4) but remain statistically significant. The weakening relationship between bank lending and physical distance is mainly caused by including social connectedness.[13] Consequently, the physical distance effect in the literature can partly be explained by the structure of social ties and, as analyzed in more detail below, the information flowing along these ties. Transportation costs, on the other hand, may still play a role as bank lending significantly decreases with physical distance even when accounting for social connectedness and cultural distance. In Section 5, we exploit explicit measures of transportation costs and find additional support for the distinct relationship between loan allocations and social connectedness.

The increase in bank lending associated with closer social ties is sizable. As the standardized beta coefficients reported at the bottom of Table 4 illustrate, a standard-deviation increase in social connectedness is associated with an increase in loan volumes by 0.12 standard deviations. The standardized beta coefficients of physical distance and cultural distance equal -0.17 and -0.06, respectively. Hence, loan volumes increase with social connectedness twice as much as they decrease with cultural differences and almost as much as they decrease with physical distance. Put differently, at a median physical distance, a standard-deviation increase in social connectedness can compensate for the lending barriers of 621 additional miles between borrower and lender, which equals 1.5 times the median distance in our sample.[14] Overall, our baseline estimates reveal the important role of social connectedness in bank lending, which is distinct from physical and cultural distances, and compensates for the lending barriers posed by these distances.[15]

– Table 4 around here –

[13] Adding social connectedness to the specification in column 2 of Table 4 changes the coefficient of physical distance from -0.389 to -0.297; adding cultural distance instead of social connectedness changes the coefficient only to -0.346 (full regressions not reported).

[14] A standard-deviation increase in social connectedness leads to an increase of log loan volumes by 0.245. This increase compensates for an increase in the *logarithm* of physical distance by 0.92. As physical distance enters the regressions in logs, the effect of physical distance is nonlinear. At the median physical distance (413 miles), the increase in the logarithm of physical distance corresponds to an increase in physical distance of 621 miles.

[15] The results hold within and across states and thus cannot be explained by changes in regulation at state borders (see Table OA5 in the Online Appendix).

Alternative specifications and the probability of bank lending As discussed in Section 2.1, connectedness and distances are slow-moving or time-invariant, which is why we obtain our baseline estimates based on cross-sectional data. However, the relationship between loan allocations and social connectedness is not limited to this cross section. For the regressions reported in columns 1 and 2 of Table 5, we exploit a panel dataset that covers the years 2004 to 2018. In column 1, we re-estimate our baseline regression (compare Table 4, column 4). The results are robust. More importantly, the same applies to the results in column 2, where we include source-county-time and destination-county-time fixed effects. We can thus exclude that our results are driven by county-time-specific credit demand or the economic development in the bank's home county. When estimating time-varying coefficients of social connectedness by interacting the variable with time dummies, loan volumes significantly increase during each of the 15 years and the size of the increase is stable over time (see Figure OA1 in the Online Appendix). The panel estimates thus strongly support that cross-county lending increases with social connectedness.

Next, we move back to cross-sectional data, but include also those county pairs that do not experience any lending (refer to the discussion toward the end of Section 3.1). Again our results are maintained (column 5). More interestingly, including all county pairs allows us to analyze the relationship between social connectedness and the probability that any lending takes place at all. Column 6 reports estimates of a linear probability model in which the regression specification deviates from our baseline regression only in that the dependent variable now is a dummy variable that equals 1 if there is at least some lending from source to destination county. The lending probability significantly increases with social connectedness, as the coefficient is positive and statistically significant. Specifically, an increase in social connectedness by one percentage point is associated with a 0.02 percentage point higher lending probability, which corresponds to a 3% increase relative to the average probability of a lending relationship. The results also hold when restricting the sample to counties that are less far apart, such that lending is more likely to take place (see Table OA6 in the Online Appendix). Hence, social connectedness increases cross-county lending both at the intensive and at the extensive margin.

– Table 5 around here –

Instrumental variable approaches As illustrated above, bank lending increases with social connectedness across a broad set of specifications, which include a variety of fixed effects. We also control for a large set of variables that may drive bank lending and are related to social connectedness, such as migration, trade, and commuting. Nevertheless, our estimates are not based on a natural experiment. In this section, we propose four instrumental variable approaches to address potentially remaining endogeneity concerns. The first three instruments are based on a historical travel cost argument. These costs do not have a direct effect on bank lending today but may have shaped social ties in the past, some of which may have persisted for generations. Compared to our baseline specification, we additionally control for present-day travel costs, such that our results cannot be explained by a correlation between historical and current travel costs. The fourth instrument exploits the quasi-random staggered introduction of Facebook across the United States, which was again not causally related to bank lending but may have shaped social connectedness, as Facebook offers a way to stay in touch.

For our first instrument, we obtain data on county-to-county highway connections from Baum-Snow (2007), who also provides highway construction dates. U.S. highways were planned during World War II to improve logistics for the war efforts and were built in the aftermath of the war, partly to facilitate a quick relocation of resources during the Cold War. While the founding fathers of the highway network were thus not motivated by considerations related to bank lending, it is conceivable that social ties have emerged along highways and that these historical social ties are persistent. To exploit this idea, we define an indicator variable that equals 1 whenever two counties are connected by the same highway and use this variable as an instrument for social connectedness. Column 1 of Table 6 reports the results. As expected, social connectedness is higher for counties that are connected by the same highway (see coefficient of the first-stage regression at the bottom of the table).[16] A test for the significance of the coefficient returns an F-value of 53, providing no indication of a weak instrument. In the second-stage regression, the coefficient of social connectedness is positive and statistically significant. The coefficients of physical and cultural distance remain negative but culture does not enter the regression significantly anymore. Hence,

[16] Table OA7 reports the full first-stage regression results.

the results support our findings on social connectedness and emphasize its comparably large effect on bank lending.

While social ties indeed appear to have emerged along highways, they likely did so slowly over time. To incorporate this idea, our next instrument measures the number of years that have passed since the construction of a highway that connects two counties. As can be seen in column 2 of Table 6, social connectedness is larger the longer a highway connection existed and the F-test rejects the presence of a weak instrument with an F-value of 48. The second-stage regression estimates indicate that a one-point increase in the social connectedness index leads to a 3.3% increase in the loan volume, which is almost identical to the results from our first instrument. We also assess the robustness of our first two instrumental-variable regressions by excluding those counties that the highways were primarily meant to connect. Specifically, we exclude particularly urban counties (i.e., beyond the 75th percentile of the distribution). The results are robust (see Table OA8, columns 1 and 2).

For our third instrument, we obtain data from Donaldson and Hornbeck (2016), who calculate travel costs for the time after the Westward Expansion, that is, the late 19th and early 20th century. The county-to-county costs are computed as the cheapest combination of traveling by railways, canals, and cattle paths. We use the latest available data (1920), as connectedness patterns were less persistent while the railway network was still under construction.[17] Similar to our first two instruments, the historical travel costs significantly correlate with social connectedness in the first stage and we find no indication of a weak instrument (F-value equals 142). In the second-stage regression, the coefficients of physical and cultural distance are again negative but insignificant, whereas social connectedness has a strong positive effect on loan volumes. Overall, all three instruments based on historical travel cost arguments emphasize that social connectedness significantly increases cross-county lending.

As an alternative approach, we exploit the quasi-random staggered introduction of Facebook as an instrument. Created by Mark Zuckerberg and his colleagues in the dorm rooms of Harvard

[17] The F-value of the test for significance of the instrument in the first-stage regression decreases the further we go back in time. We can go back as far as 1880 before the instrument becomes weak. Until then, the results are independent of the choice of year (regressions not reported).

University, Facebook memberships were initially restricted to students from this university.[18] Later on, the online network was gradually opened to other Ivy League colleges and, afterward, in a quasi-random fashion to other universities. Due to the initial pattern, we subsequently exclude counties that host Ivy League colleges. We track how Facebook spread across the United States by manually recovering the order in which the first student of a university created a Facebook account.[19] We combine this hand-collected information with university locations and rank counties by the appearance of Facebook in these counties.[20] The instrument is then defined as follows:

$$Facebook\ rollout = \frac{Rank_i + Rank_j}{Student\ population_i + Student\ population_j}\ . \tag{6}$$

$Rank_i$ ($Rank_j$) is the rank number of county i (j). We scale this sum by the student populations at the time of the rollout (i.e., 2005) to account for the possibility that universities with more students joined Facebook earlier simply because of the larger number of students.

Column 4 of Table 6 displays the results based on the Facebook rollout as an instrument.[21] The instrument significantly correlates with social connectedness (the F-value equals 163). According to the second-stage estimates and in line with our other instrumental variable approaches, social connectedness significantly increases loan volumes. Overall, the results thus strongly emphasize social connectedness as a key driver of loan allocations.

– Table 6 around here –

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

[19] 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 which universities early Facebook users studied at, this information enables us to recover the order in which universities gained access to Facebook.

[20] In some regions, Facebook only took off after the construction of user IDs had been randomized. We set the rank of these late joiners to the maximum value of the rank distribution plus one standard deviation.

[21] In this regression, we use an absolute measure of social connectedness to avoid a mechanical correlation between the denominator of the instrument (student populations) and the denominator of the social connectedness index (total populations) and, instead, add the population product as an additional control variable. The results are qualitatively robust when we follow our usual specification (see Table OA8, column 4). However, in that case, the second-stage coefficient of social connectedness appears inflated, such that we believe it to be prudent to emphasize our main instrumental-variable results in Table 6.

4.1.2 Information sensitivity of lending

The previous findings are consistent with social ties facilitating banks' access to soft information. Subsequently, we explore the information channel more closely by analyzing how the effect of social connectedness depends on the information sensitivity of loans.

Banks, borrowers, and the economic environment Lending processes tend to be less standardized in small banks, which leaves more room for soft information to feed into lending decisions. Based on this idea, we begin our analysis of the information channel by adding an interaction term between social connectedness and bank size to our baseline regression (compare Equation 1). Bank size is defined as the logarithm of the loan-volume-weighted average total assets of all banks that lend from source to destination county. Table 7, column 1, reports the results. Social connectedness indeed increases loan volumes more strongly for smaller banks. Based on this regression, a plot of the effect of social connectedness at different levels of bank size shows that the effect of social connectedness is three times as large as our baseline estimate for county pairs that predominantly experience lending from small banks, whereas the effect becomes just insignificant for the very largest banks (Figure OA2, panel (a)). In column 2, we interact social connectedness with the borrower county's exposure to industry volatility. To calculate this variable, we weight the U.S.-wide output volatility of industries with a county's industry shares. Banks are likely to find it more difficult to judge a small or medium-sized enterprise's ability to repay if they expect this firm to operate in a more unpredictable economic environment. Our results indicate that the effect of social connectedness is larger for those more opaque borrowers, as the coefficient of the interaction term is positive and statistically significant while the coefficient of social connectedness remains positive and significant. The coefficients imply that loan volumes to counties that are exposed to volatile industries (e.g., agriculture, forestry, fishing, and hunting) increase 50% more than loan volumes to counties that are exposed to more stable industries (e.g., educational services).[22] We

[22] To be precise, this number applies for a county that is exposed only to agriculture, forestry, fishing, and hunting (volatility exposure = 6.57) compared to a county that is only exposed to educational services (2.96) or any convex combinations of industries with the same volatility exposures.

thus find preliminary evidence that social connectedness affects lending decisions more strongly if banks are more in need of information.

Next, we interact social connectedness with deciles of GDP growth in the borrower's county. If the effect of social connectedness is related to loan officers lending to peers who struggle to obtain funding, we would expect social connectedness to increase loan volumes especially in counties where economic conditions are weak. Our estimates, however, show that the effect of social connectedness is stronger both in borrower counties that experience particularly low and those that experience particularly high GDP growth. The coefficient of social connectedness, which represents the effect of social connectedness at the 5th decile of borrower-county GDP growth, is positive and statistically significant, as are the interactions between social connectedness and the first two and the last decile (column 3). Providing further evidence of the information channel, we thus find evidence that social connectedness increases lending most strongly when banks are confronted with an unusual local economic development such as a strong boom that might signal unsustainable growth or a strong bust, e.g. related to a larger firm that moves its production away from the borrower county.

Lastly, banks may have a higher need for information when lending to an economic environment that develops differently from their home market. To exploit this idea, we interact social connectedness with deciles of the GDP growth differential between source and destination county. The coefficient of social connectedness and its interaction with the four highest deciles of the GDP growth differential are positive and significant, whereas all other interactions do not enter the regression significantly (column 4). Hence, bank lending increases more strongly in social connectedness when the local economic environment of the borrower and the lender are particularly different, which again illustrates that social connectedness has stronger effects if the banks' need for information is high.

– Table 7 around here –

Loan types To further explore the role of information, we subsequently distinguish between types of loans that differ in their screening incentives. The analysis is based on the mortgage-lending data, which allow the identification of different loan types. In our baseline specification, the effect of social connectedness is smaller in the mortgage loan sample than in the SME sample (see Table OA9 and the standardized beta coefficients therein). This also points to an information channel, as the credit intermediation process tends to be less standardized for SME loans and small and medium-sized enterprises are more opaque borrowers, such that soft information is more important for SME lending than for mortgage lending.

Our first distinction between mortgage loan types is based on government guarantees. These guarantees protect banks from default risk, which reduces screening incentives and, hence, makes information less valuable. We aggregate cross-county mortgage loans with and without guarantees separately and run our baseline regression on these two different subsets. Table 8, column 1, displays the results for loans which are *not* backed by government guarantees. For these loans, the coefficient of social connectedness is highly significant and equals 0.016, which is almost twice as large as in the overall sample (0.009, compare Table OA9, column 2). Conversely, social connectedness is not significantly related to lending for loans that are backed by government guarantees (column 2 of Table 8). We formally test if the effect of social connectedness significantly differs across the two loan types by adding an interaction term between social connectedness and the share of the volume of guaranteed loans to our baseline regression (column 3). In line with our previous findings, loan volumes significantly increase with social connectedness, but the effect becomes significantly smaller the higher the guaranteed share. Supporting our earlier reasoning, social networks thus play an important role in bank lending if banks bear the risk of a loan, thus having an incentive to screen and to make use of the information flowing through these networks.

As an additional source of variation in the information sensitivity of lending, we next distinguish between loans that are kept on the originating bank's balance sheet and loans that are securitized. Banks reduce screening activities for securitized loans but the incentives to screen are not entirely eliminated because of reputation concerns (Keys, Mukherjee, Seru, and Vig, 2010; Purnanandam, 2011; Keys, Seru, and Vig, 2012; Wang and Xia, 2014).[23] Hence, access to information through social networks should be less relevant for these loans. While the coefficient of social connectedness is positive and statistically significant in both samples, loan volumes increase twice

[23]For a discussion of agency conflicts in securitization see, for instance, Fenner, Klein, and Mössinger (2019).

as much in social connectedness for loans that are kept on the books compared to loans that are securitized (Table 8, columns 4 and 5). Again, we test whether the difference in the effect of social connectedness across loan types is statistically significant based on an interaction between social connectedness and the share of the volume of securitized loans (column 6). The estimates support the existence of the differential effect.

– Table 8 around here –

Overall, the results in this section demonstrate that social connectedness increases cross-county lending, especially if banks have a high need for information and strong screening incentives. The findings thus strongly support that social connectedness plays an important role in bank lending *because* banks can leverage their social network as a source of information.

4.2 Riskiness of loans

Above, we have analyzed how social connectedness affects which counties banks lend to. Subsequently, we explore whether the type of borrowers that receive loans (risky vs. less risky), the loan conditions (interest rate), and the performance of loans also differs across social connectedness. This allows us to further investigate the information channel and to assess consequences of social connectedness altering banks' lending decisions. The analysis exploits our loan-level sample of mortgage loans (see Section 3.2). Table 9 reports the results.

In column 1, we regress the borrower's FICO score at the time of origination of the loan on social connectedness, while controlling for physical and cultural distance, and bank and origination-year fixed effects (see Equation 2 in Section 2.2). The coefficient of social connectedness is insignificant. Borrowers' creditworthiness thus is not heterogeneous across social connectedness. In column 2, we estimate the same model to explain loan-to-value (LTV) ratios. The coefficient of social connectedness is again not statistically significant. From an ex-ante perspective, the riskiness of a loan is thus independent of the social ties between a borrower's and a bank's regions.

In column 3, we use the interest rate (in basis points) as the dependent variable while additionally controlling for the FICO score and the LTV ratio, the loan volume (in logs), the debt-to-income ratio, and a binary variable that indicates first-time home buyers. In line with expectations, the estimates indicate that the interest rate decreases with a borrower's creditworthiness but increases with their LTV ratio. More importantly, the coefficient of social connectedness is negative and statistically significant. According to our estimates, borrowers with a standard deviation higher social connectedness pay a 2.7 basis points lower interest rate (=40*(-0.068)), which equals 3% of a standard deviation of the interest rate (83). Hence, borrowers from well-connected counties not only receive more lending, but they also have access to cheaper financing.

To analyze how social connectedness relates to the ex-post loan performance, our next dependent variable indicates delinquent loans. The variable equals 1 if a loan has been at least 90 days past due at least once. The estimates reveal no statistically significant relationship between the probability of delinquency and social connectedness (column 4). In column 5, we focus on more extreme cases, namely loans that actually default. Specifically, we regress the unpaid balance on social connectedness and an interaction of social connectedness and a dummy variable that equals 1 if a loan is *not* in default. The coefficient of social connectedness is negative and statistically significant. Its sum with the interaction term is insignificant. Hence, controlling for the original loan amount, the origination year, the ex-ante riskiness, and further loan characteristics, the remaining loan amount significantly decreases with social connectedness *for those loans that are in default.* Specifically, the estimates imply that borrowers who default owe the bank 80% less if they are from a region with one standard deviation higher social connectedness. Banks thus profit from superior performance of loans to well-connected regions.

– Table 9 around here –

Overall, the loan-level analysis shows that social connectedness is not associated with lending to riskier borrowers. However, borrowers from well-connected counties pay lower interest rates. This result is in line with both a lower cost of information acquisition for banks and their expectation of improved loan performance due to access to superior information. We indeed find evidence of improved performance of loans. While delinquency rates do not differ across social connectedness, defaulting loans cause much smaller losses if social connectedness is high. Since social connectedness is not associated with an increased ex-ante riskiness of lending, the results are in stark contrast to the effects of a preferential treatment of peers that results in the financing of negative NPV projects. Instead, the results are in line with social ties facilitating banks' access to information, which can benefit both borrowers and banks.

4.3 Real effects

To further explore the consequences of the role of social connectedness in bank lending, we subsequently analyze the real effects of borrower counties' social proximity to bank capital. The analysis is based on our county-level dataset, for which we aggregate the county-pair-level social connectedness at the county level by calculating a borrower county's social proximity as its average social connectedness to all counties weighted by total bank assets in these counties (see Section 3.3). Table 10 reports the results of our analysis.

In column 1, we regress the volume of SME loans to the borrower county (in logs) on social proximity, while controlling for physical and cultural proximity, county- and state-time fixed effects, industry shares, commuting, and migration (see Equation 3 in Section 2.3). The coefficient of social proximity is positive and statistically significant. It indicates that a standard-deviation increase in a county's social proximity to bank capital increases the total volume of SME lending to that county by 4.7%. In line with our earlier findings, borrowers from regions with closer social ties to banks' regions receive more lending.

In column 2, we re-estimate our model with real GDP growth as the dependent variable. The coefficient of social proximity is positive and highly significant. According to our estimates, counties with one standard deviation higher social proximity to bank capital experience 0.85 percentage points higher GDP growth. Note that *county*-level GDP growth generally fluctuates more than *country*-level growth. The 0.85 percentage points increase equals an increase of 11% (=0.85/7.6) of a standard deviation of GDP growth, which is sizable and well in line with a long list of studies of the real effects of access to bank funding.[24] In column 3, we additionally interact social proximity with the percentage of small firms in the borrower county. Small firms tend to rely more on bank lending

[24]Starting from Jayaratne and Strahan (1996) and Levine and Zervos (1998) to recent studies such as Huber (2018).

for financing. The coefficient of social connectedness and the coefficient of its interaction with the small-firm percentage are positive and statistically significant. Hence, GDP growth increases more strongly with social proximity in counties with many small firms.

Similar findings hold for employment. In columns 4 and 5 we use the number of employed people (in logs) as dependent variable. Employment significantly increases with social connectedness (column 4). A standard-deviation increase in a county's social proximity is associated with a 0.4% increase in employment. This increase is again larger for counties with a higher percentage of small firms, as indicated by the positive and significant interaction term in column 5. The size of the increase is again in line with the literature.[25]

– Table 10 around here –

Overall, counties with higher social proximity to bank capital receive more lending and have higher GDP growth and more employment. We thus find strong additional evidence that borrowers profit from strong social ties between their own region and a bank's region.

5 Additional analyses and robustness checks

This section provides two types of additional analyses that complement our findings on cross-county lending. First, it analyzes how the effects of connectedness and distances depend on each other. Second, it assesses the robustness of our baseline results with respect to alternative measures of physical and cultural distance. Thereby, this section reaffirms that our results are independent of the chosen measurement approaches and that the effects of social connectedness are distinct from physical and cultural distances.

5.1 Social connectedness and loan allocations: heterogeneities across distances

The findings in Section 4.1.1 clearly illustrate the potential of social ties to compensate for the lending barriers posed by physical and cultural distance. This section explores nonlinear effects of

[25]For instance, Huber (2018) estimates that a standard-deviation increase in a county's dependence on a weakly capitalized major bank reduces county-level employment by 0.83% in Germany.

social connectedness. In so doing, it also takes the analysis one step further and discusses whether the lending barriers associated with distances disappear in the case of sufficiently close social ties.

Table 11 displays the results. For ease of comparability, column 1 restates our baseline regression from Table 4, column 4. We begin the discussion of nonlinear effects by analyzing whether the effect of social connectedness depends on its level. To this end, we add a squared term of this variable to our baseline regression. Our baseline estimates remain unchanged and the coefficient of the squared term is not statistically significant (column 2). Hence, loan volumes increase linearly in social connectedness.

More interestingly, we explore whether social ties become more or less important for lending decisions as the physical distance between borrower and lender increases. The results suggest the latter. Whereas the coefficient of social connectedness remains significantly positive, the coefficient of its interaction with physical distance is negative and statistically significant (column 3). As distance increases, the positive effect of social connectedness on loan volumes decreases. We conjecture that this decreasing effect is associated with a decreasing *intensity* of social ties (as opposed to their number) at larger distances, where opportunities for face-to-face contact are more rare. As a result, distant contacts exchange less information, making them less valuable for bank-lending decisions.

Whereas the effect of social connectedness decreases with physical distance, it increases with cultural distance. When extending our baseline specification by an interaction term between social connectedness and cultural distance, its coefficient is positive and significant (column 4). The coefficient of social connectedness remains significantly positive. Hence, loan volumes increase with social connectedness, but this relationship is particularly pronounced at large cultural distances. In fact, the negative effect of cultural distance disappears entirely in the case of sufficiently close social ties in our sample. Social connections thus bridge a cultural divide between borrower and lender. The channel for this effect can be twofold. First, the information flowing through social networks may reduce statistical discrimination that emerges if loan officers with one cultural background lack the information to fully assess loan applicants from a differing cultural background. Second, strong social ties may overcome discrimination due to (subconscious) prejudices against people of unfamiliar cultural backgrounds. In both ways, social connectedness may reduce the lending barrier posed by cultural differences.

– Table 11 around here –

5.2 Alternative measures of physical distance

In our baseline regressions, physical distance is measured as the great-circle distance (i.e., "as the crow flies") between two counties, where county locations are based on the geographical center of a county. If two neighboring counties each host a city close to their common border, social connectedness and cross-county lending may both be increased due to the lower physical distance. This, however, would not be reflected by our measure. While we control for counties that share a common border in all our regressions, we subsequently assess this alternative explanation for our results more closely. Specifically, we define physical distance as the great-circle distance between *population-weighted* county centroids. Column 1 of Table 12 restates our baseline results from Table 4, column 4. When employing the alternative definition of physical distance, the coefficients of social connectedness and cultural distance remain identical, while the coefficient of physical distance decreases slightly but remains statistically highly significant (column 2). Our results thus cannot be explained by an imprecise identification of cities that are located close to county borders.

A similar argument applies to the infrastructure between borrower and lender. For example, a more convenient road connection between borrower and lender may simultaneously increase social connectedness and lending, which our measure of physical distance cannot fully account for. To assess this hypothetical explanation for our findings, we use the data on county-to-county road travel costs from the Oak Ridge National Laboratory's National Transportation Center as an alternative measure of physical distance. Once again, the results are unchanged as the coefficients of social connectedness, physical distance, and cultural distance are almost identical to our baseline estimates (column 3). The same applies when accounting for the cheapest combination of road, railway, and waterway travel (column 4). Consequently, our results are robust across measures of physical distance, which supports that the effect of social connectedness is distinct from physical distance.

– Table 12 around here –

5.3 Alternative measures of cultural differences

To construct our measure of cultural distance, we combine information on ethnic ancestries, racial origins, religious beliefs, and the structure of people's social environment into a single variable. This variable has the advantage of being interpretable as the cultural distance between two counties. However, the aggregation requires us to weight the underlying information (see Section 3.1). Subsequently, we explore the robustness of our results with respect to the method of aggregation and the overall measurement approach.

Table 13 reports the results. Column 1 restates our baseline regression from Table 4, column 4. In column 2, we include the four dimensions of cultural identity separately. The effect of social connectedness remains unchanged and loan volumes decrease only slightly more in physical distance compared to our baseline estimates. The sign of the coefficients of all four cultural variables is negative, indicating that all aspects of cultural differences hamper bank lending. However, only the coefficient of the social environment is statistically significant. When we exclude this variable in column 3, the coefficients of the three remaining culture variables are negative, but racial origin now enters the regression significantly. All other results remain unchanged. We can thus exclude that our findings on culture or any other findings are driven by the social-environment dimension in our culture data. In column 4, we include all 39 culture variables separately, without any change to our findings. Consequently, our results are unaffected by how we use the information on cultural backgrounds to account for the cultural distance.

Next, we measure cultural differences based on a different approach. More precisely, we proxy cultural differences with the county-pair-specific vote-share differential for the Republican candidate during the 2016 presidential election, as voting patterns are partially an outcome of cultural patterns (see, for example, Lieske (1993)). Column 5 reports the results for our baseline regression when applying this alternative measure of cultural distance. 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 – relevant for lending outcomes even in a within-country setting. Importantly, when using this alternative approach to the measurement of cultural distance, the coefficients of social connectedness and physical distance remain largely unchanged.

Lastly, we explore whether the relationship between bank lending and social connectedness depends on the cultural heterogeneity in the destination county. To this end, we collect our culture data on the census tract level and calculate the cultural distance between census tracts within a county in the same way in which we calculated the cultural distance between counties.[26] We then calculate the average cultural distance between census tracts per county and sum these averages for each county pair. While the coefficient of social connectedness remains positive and significant, the interaction term between cultural heterogeneity and social connectedness is insignificant. Therefore, our findings cannot be explained by social connectedness serving as a proxy for within-county cultural heterogeneity. The results thus provide additional evidence that the relationship between social connectedness and bank lending is distinct from cultural distance.

– Table 13 around here –

6 Conclusion

This paper analyzes how the geographic structure of social ties affects bank lending. While the previous literature directs its attention toward the lending between peers in exclusive networks or the physical and cultural distance between borrower and lender, we focus on the ubiquitous social network that spans a society. This network can help to overcome asymmetric information by facilitating banks' access to information about borrowers or their local economic environments without requiring a direct link between loan officers and borrowers.

We find that banks from one region lend more to another region if the people who live in these two regions are more connected. This effect of social connectedness is large and compensates for the lending barriers posed by physical and cultural distances. Social connectedness increases bank lending particularly strongly if banks have a high need for information and screening incentives are intact. At the same time, social connectedness does not result in lending to riskier borrowers

[26] We have to leave out the data on religious backgrounds, because it is unavailable at this granular level.

but is associated with lower borrowing costs and improved loan performance. In addition to higher lending, counties with higher social proximity to bank capital exhibit higher GDP growth and more employment. Consequently, banks and especially borrowers profit from social connectedness, which affects loan allocations because of the information that moves along social ties.

While our results primarily reveal the important role of social connectedness as an information channel in bank lending, they have several potential implications that may also be of interest for future research. Antitrust policies may benefit from taking the structure of social networks into account. Social connectedness increases lending and partly explains the effect of physical distance, meaning that a high concentration of lenders in an area is less concerning if banks outside of this area are well connected to it. Second, social connectedness may help to explain the trend toward geographically more dispersed banking, as banks obtain information through social networks which themselves have become increasingly widespread. Lastly, banks may reduce information asymmetries by employing well-connected agents to obtain information, especially when attempting to expand business in culturally different regions, as the lending barriers posed by different cultural backgrounds disappear when social ties are sufficiently close.

References

Agarwal, S., and R. Hauswald (2010): "Distance and private information in lending," *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., P. Farrell, T. Kuchler, and J. Stroebel (2019): "Social connectedness in urban areas," *Working Paper*.

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., S. Ongena, and İ. Şendeniz-Yüncü (2019): "Keep walking? Geographical proximity, religion, and relationship banking," *Journal of Corporate Finance*, 55, 49–68.

Berger, A. N., N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. 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*, 9(1), 7–25.

Brevoort, K. P., and T. H. Hannan (2006): "Commercial lending and distance: evidence from Community Reinvestment Act data," *Journal of Money, Credit and Banking*, 38(8), 1991–2012.

Brown, J. R., Z. Ivković, P. A. Smith, and S. J. Weisbenner (2008): "Neighbors matter: Causal community effects and stock market participation," *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," *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," *Journal of Finance*, 65(4), 1409–1438.

Degryse, H., L. Laeven, and S. Ongena (2009): "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," *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.

Diamond, D. W. (1984): "Financial intermediation and delegated monitoring," *Review of Economic Studies*, 51(3), 393–414.

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," *Quarterly Journal of Economics*, 118(3), 815–842.

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

Fenner, A., P. Klein, and C. Mössinger (2019): "Digging into the black box of portfolio replenishment in securitization: Evidence from the ABS loan-level initiative," *Working Paper*.

Fisman, R., D. Paravisini, and V. Vig (2017): "Cultural proximity and loan outcomes," *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," *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 and Finance*, 87 (2018), 26–39.

Halim, E., Y. Riyanto, and N. Roy (2019): "Costly information acquisition, social networks, and asset prices: Experimental evidence," *Journal of Finance*, 74(4), 1975–2010.

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.

Hollander, S., and A. Verriest (2016): "Bridging the gap: The design of bank loan contracts and distance," *Journal of Financial Economics*, 119(2), 399–419.

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

——— (2005): "Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers," *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.

Huber, K. (2018): "Disentangling the effects of a banking crisis: Evidence from German firms and counties," *American Economic Review*, 108(3), 868–898.

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," *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.

Jayaratne, J., and P. E. Strahan (1996): "The finance-growth nexus: Evidence from bank branch deregulation," *Quarterly Journal of Economics*, 111(3), 639–670.

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

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010): "Did securitization lead to lax screening? Evidence from subprime loans," *Quarterly Journal of Economics*, 125(1), 307–362.

Keys, B. J., A. Seru, and V. Vig (2012): "Lender screening and the role of securitization: Evidence from prime and subprime mortgage markets," *Review of Financial Studies*, 25(7), 2071–2108.

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

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

Kuchler, T., Y. Li, L. Peng, J. Stroebel, and D. Zhou (2020): "Social proximity to capital: Implications for investors and firms," *NBER Working Paper (27299)*.

Kuchler, T., D. Russel, and J. Stroebel (2020): "The geographic spread of COVID-19 correlates with structure of social networks as measured by Facebook," *Working Paper*.

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," *Journal of Finance*, 57(1), 265–301.

Levine, R., and S. Zervos (1998): "Stock markets, banks, and economic growth," *American Economic Review*, 88(3), 537–558.

Liberti, J. M. (2018): "Initiative, incentives, and soft information," *Management Science*, 64(8), 3714–3734.

Lieske, J. (1993): "Regional subcultures of the United States," *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," *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.

Nguyen, H.-L. Q. (2019): "Are credit markets still local? Evidence from bank branch closings," *American Economic Journal: Applied Economics*, 11(1), 1–32.

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

Purnanandam, A. (2011): "Originate-to-distribute model and the subprime mortgage crisis," *Review of Financial Studies*, 24(6), 1881–1915.

Saadi, V. (2020): "Role of the community reinvestment act in mortgage supply and the U.S. housing boom," *Review of Financial Studies*, forthcoming.

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

Wang, Y., and H. Xia (2014): "Do lenders still monitor when they can securitize loans?," *Review of Financial Studies*, 27(8), 2354–2391.

Figures and Tables

Figure 1: An example of a county's social connectedness

The figure displays the social connectedness between Montgomery County, Ohio, and all other U.S. counties. Dark-blue-colored areas exhibit the highest social connectedness, light colored areas the lowest social connectedness. Montgomery County is represented by the white spot in the middle of the cluster of dark-blue counties. The county is representative of our sample in terms of the correlation between social connectedness, physical distance, and cultural distance.

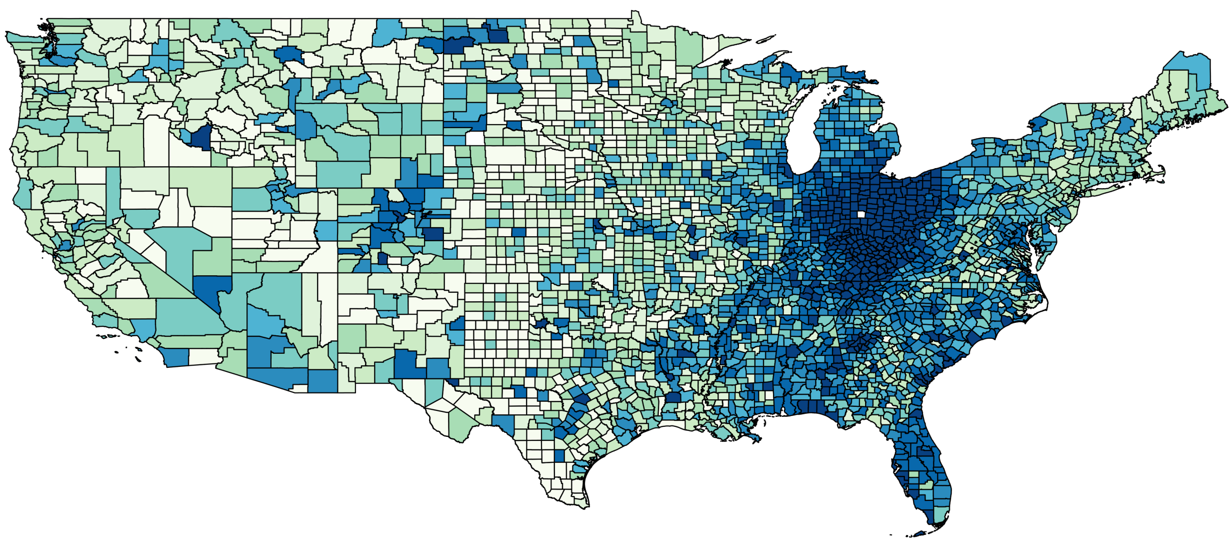

Figure 2: Clusters of regional subcultures in the United States

The figure displays regional subcultures in the United States based on a cluster analysis of our culture data. For this analysis, we determine the principal components of our 39 individual culture variables. We keep the ten principal components with eigenvalues larger than 1 and sort counties into seven clusters by minimizing the mean of the Euclidean distance between the principal components' scores. The number of clusters is chosen based on an elbow method with 1,000 repetitions of randomly chosen starting counties for each cluster. The figure's pattern is robust to the number of principal components and clusters.

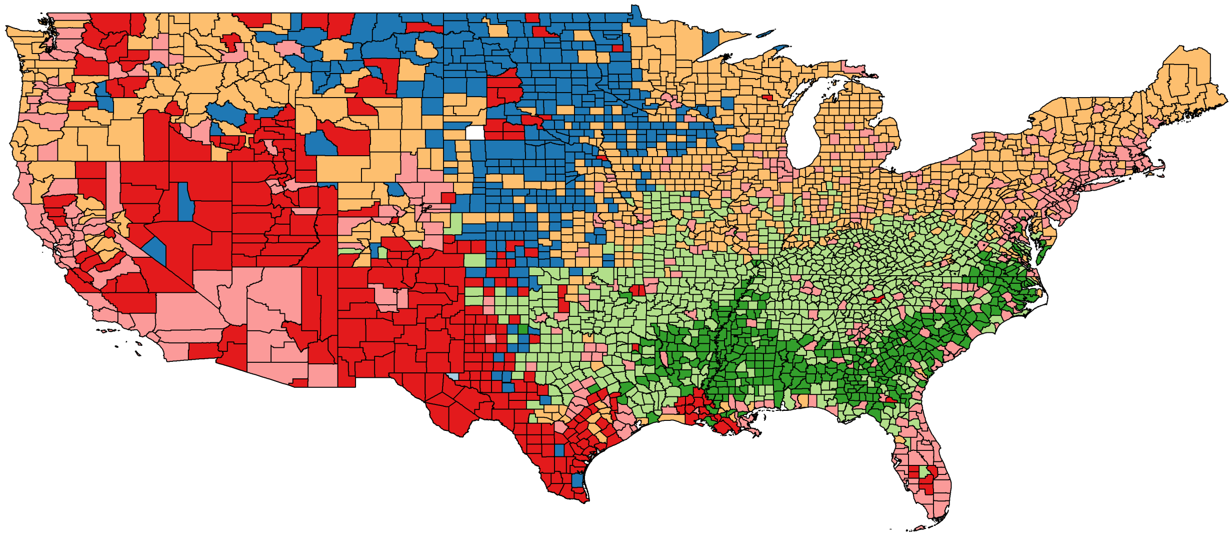

Figure 3: Counties' social proximity to bank capital

The figure illustrates each county's social proximity to bank capital. Dark-blue-colored areas exhibit the highest social proximity, light-colored areas the lowest social proximity. County i's social proximity to bank capital at time t is defined as $\sum_j Social\ connectedness_{i,j} \cdot Total\ assets_{j,t}$. The figure displays averages over the years 2009 to 2017.

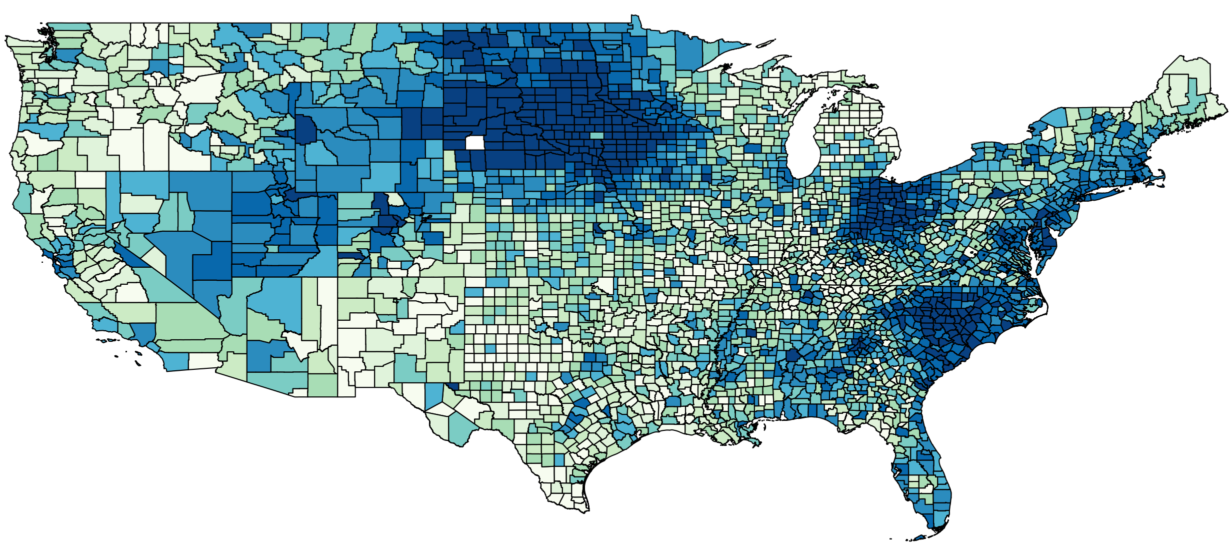

Table 1: Descriptive statistics

The table displays descriptive statistics for the main sample used to analyze social connectedness and the allocation of bank lending. All variables are at the county-pair level. For instance, "volume of SME loans" equals the volume of loans provided by all bank branches in a source county to the small and medium-sized enterprises in the destination county. Table A1 in the appendix summarizes variable definitions and data sources. Tables 2, 3, and OA3 report descriptive statistics for the additional samples used to analyze heterogeneities across the information sensitivity of loans in Section 4.1.2, the riskiness of loans in Section 4.2, and real effects in Section 4.3.

	N	Mean	Median	SD	Min	Max
Connectedness and distances						
Social connectedness	66,684	20	2	35	0	100
Cultural distance	66,684	17	16	7	0	72
Physical distance [miles]	66,684	578	413	566	4	4,996
log(Physical distance)	66,684	5.8	6.0	1.3	1.5	8.5
Lending data						
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
County-pair-level control variables						
Common border	66,684	0.1	0	0.3	0	1
GDP growth differential [ppt.]	66,684	3.6	2.7	3.4	0.0	46.9
Gross commuting [%]	66,684	0.1	0.0	0.6	0.0	16.9
Gross migration [%]	66,684	0.03	0.00	0.13	0.00	3.20
Gross trade [million USD]	66,684	85	38	114	0	814
Same state	66,684	0.2	0	0.4	0	1
Unemployment differential [ppt.]	66,684	1.5	1.1	1.4	0.0	21.4
Instrumental variables						
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
Same highway	66,684	0.1	0	0.3	0	1
Years since highway construction	66,684	5	0	15	0	79
Further variables from interaction terms						
Bank size	66,684	17.3	17.5	2.1	0.0	21.5
Borrower's volatility exposure	66,684	2.4	2.2	0.9	0.5	9.3
GDP growth	66,684	1.8	1.5	4.0	-26.7	47.3

Table 2: Descriptive statistics: loan-level sample of borrower riskiness and loan performance

The table displays descriptive statistics for the loan-level sample used to analyze the relationship between social connectedness and the ex-ante and ex-post riskiness of loans. "Combined LTV" is the combined loan-to-value ratio of all mortgages on the borrower's property. "DTI" is the borrower's debt-to-income ratio. "FICO score" is a measure of the borrower's creditworthiness. "Delinquent" is a dummy variable which indicates mortgages that have been at least 90 days past due at least once. Table A1 provides variable definitions and data sources.

	N	Mean	Median	SD	Min	Max
Connectedness and distances						
Social connectedness	20,760	26	2	40	0	100
Cultural distance	20,760	16	16	7	1	44
Physical distance [miles]	20,760	551	420	521	3	4,152
log(Physical distance)	20,760	5.6	6.0	1.4	1.4	8.3
Ex-ante loan risk						
Combined LTV	20,760	82	80	15	5	120
DTI	20,760	35	35	12	2	65
FICO score	20,760	727	737	56	300	835
Loan characteristics						
First-time buyer	20,760	0.2	0	0.4	0	1
Interest rate [basis points]	20,760	648	638	83	425	1,075
Original loan amount [thousand USD]	20,760	161	150	88	6	802
log(Original loan amount)	20,760	4.9	5.0	0.6	1.8	6.7
Ex-post loan performance						
Delinquent	20,760	0.1	0	0.2	0	1
Not in default	20,760	0.992	1	0.1	0	1
Unpaid balance [thousand USD]	20,760	79.9	38	96.8	0	753
log(Unpaid balance)	20,760	-0.4	3.6	5.8	-6.9	6.6

Table 3: Descriptive statistics: county-level sample for the analysis of real effects

The table displays descriptive statistics for the county-level analyses of the real effects of counties' social proximity to bank capital. For readability, we do not display summary statistics for the industry shares used as control variables. Table A1 provides variable definitions and data sources.

	N	Mean	Median	SD	Min	Max
Connectedness and distances						
Social proximity	24,161	1.1	0.7	1.0	0.1	8.3
Cultural proximity	24,161	4.5	4.4	1.0	1.2	9.2
Physical proximity	24,161	0.6	0.5	1.0	0.1	38.0
Dependent variables						
Employment	24,161	47,459	11,024	152,467	62	4,883,640
log(Employment)	24,161	9.5	9.3	1.5	4.1	15.4
Loan volume [thousand USD]	24,152	66,996	9,818	248,022	1	8,843,872
log(Loan volume)	24,152	16.2	16.1	1.9	6.9	22.9
Real GDP growth [%]	24,161	1.5	1.1	7.6	-20.1	33.3
Control variables						
Commuting from county	24,161	40,068	8,556	138,144	0	4,516,714
log(Commuting from county)	24,161	8.3	9.1	3.4	0.0	15.3
Commuting to county	24,161	39,903	7,167	149,760	0	4,665,782
log(Commuting to county)	24,161	8.1	8.9	3.4	0.0	15.4
Migration	24,161	6,128	1,604	15,324	1	289,585
log(Migration)	24,161	7.5	7.4	1.5	0.0	12.6
Small-firm share	24,161	57.3	56.8	12.4	8.2	100.0

Table 4: Social connectedness and the allocation of bank lending

The table provides estimates of how the allocation of cross-county bank lending varies with county pairs' social 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 between the coefficients in columns 1 to 3 and those in column 4. The standardized beta coefficients at the end of the table express the effect of a standard-deviation increase in the explanatory variable in standard deviations of the dependent variable. Control variables vary at the county-pair level and consist of the GDP growth and unemployment differentials between the two counties, gross migration and trade, the share of the commuting population, as well as same state and common border indicator variables. Table OA4 reports the coefficients of the control variables. Table A1 defines the variables. 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	0.012***			0.007***
	(0.001)			(0.001)
Physical distance		-0.389***		-0.267***
		(0.041)		(0.051)
Cultural distance			-0.034***	-0.016**
			(0.007)	(0.007)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684	66,684
Adj. R^2	0.519	0.522	0.515	0.525
Adj. R^2 within	0.121	0.126	0.114	0.132
P-value for H0: no difference to coefficient in column (4)				
Social connectedness	0.000			
Physical distance		0.063		
Cultural distance			0.057	
Standardized beta coefficients				
Social connectedness	0.21			0.12
Physical distance		-0.25		-0.17
Cultural distance			-0.13	-0.06

Table 5: Choice of sample, additional fixed effects, and the probability of bank lending

This table illustrates the robustness of our results with respect to the construction of our sample and additional fixed effects. Furthermore, it explores the relationship between social connectedness and the probability of lending between county pairs. The regressions in columns 1 and 2 are based on panel data that cover the years 2004 to 2018. Figure OA1 displays the evolution of the effect of social connectedness over time. Column 3 exploits a cross-sectional sample that also includes the county pairs that do not experience any cross-county lending. Based on this sample, column 4 analyzes the probability of a lending relationship between a source county and a destination county. Table OA6 reports additional specifications. Control variables account for 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. Table A1 defines the variables. 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)			Lending indicator
Social connectedness	0.006***	0.006***	0.003***	0.0002***
	(0.001)	(0.001)	(0.000)	(0.000)
Physical distance	-0.272***	-0.285***	-0.143***	-0.008***
	(0.052)	(0.053)	(0.013)	(0.001)
Cultural distance	-0.017***	-0.017***	-0.004***	-0.0002**
	(0.006)	(0.006)	(0.001)	(0.000)
Source county FE	Yes	No	Yes	Yes
Destination county FE	Yes	No	Yes	Yes
Source-county-time FE	No	Yes	No	No
Destination-county-time FE	No	Yes	No	No
County-pair-level control variables	Yes	Yes	Yes	Yes
No. of obs.	788,817	788,817	9,526,482	9,526,482
Adj. R^2	0.482	0.542	0.337	0.332
Adj. R^2 within	0.142	0.155	0.077	0.068

Table 6: Social connectedness and the allocation of bank lending: instrumental-variable approaches

The table provides instrumental-variable estimates for our baseline regression (see Equation 1 and Table 4, column 4). The bottom part of the table reports first-stage coefficients, p-values, and F-statistics of the instruments. Full first-stage regressions are reported in Table OA7. 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. In columns 1 to 3, we additionally control for present-day highway-travel costs. 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 OA8, column 3). Table A1 defines the variables. 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 costs	Facebook rollout
Social connectedness	0.032***	0.033***	0.024***	0.012**
	(0.011)	(0.011)	(0.005)	(0.006)
Physical distance	-0.342**	-0.333*	-0.509***	-0.187
	(0.165)	(0.172)	(0.108)	(0.118)
Cultural distance	-0.005	-0.004	-0.008	0.007
	(0.009)	(0.009)	(0.007)	(0.008)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes	Yes
Present-day travel costs	Yes	Yes	Yes	No
No. of obs.	66,647	66,647	56,223	56,852
Adj. R^2	0.496	0.494	0.520	0.539
Adj. R^2 within	0.078	0.075	0.113	0.114
Instrument (1st stage)	3.783***	0.071***	4.105***	163.1***
	(0.000)	(0.000)	(0.000)	(0.000)
F-value (1st stage)	53.3	48.1	142.1	151.2

Table 7: Information sensitivity of loans across banks and across borrowers' economic environments

This table reports regressions used to analyze how the relationship between bank lending and social connectedness depends on banks' need for information. In columns 1 and 2, we add an interaction between social connectedness and bank size or, as a measure of borrower opacity, the borrower county's demeaned exposure to industry-level volatility to our baseline regression. Columns 3 and 4 add interactions between social connectedness and dummy variables that indicate deciles of the destination-county GDP growth or the source and destination counties' GDP growth differential instead. The dependent variable is the logarithm of the volume of cross-county loans to small and medium-sized enterprises. Control variables are the physical and cultural distances, the GDP growth and unemployment differentials, gross migration and trade, the share of the commuting population, same state and common border indicator variables, and the single terms of interactions. Table A1 defines the variables. 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	0.037***	0.007***	0.006***	0.006***
	(0.006)	(0.001)	(0.001)	(0.001)
Social connectedness · Bank size	-0.002***			
	(0.000)			
Social connectedness · Borrower's volatility exposure		0.001***		
		(0.000)		
Social connectedness · GDP growth decile 1			0.002**	
			(0.001)	
Social connectedness · GDP growth decile 2			0.003***	
			(0.001)	
Social connectedness · GDP growth decile 3			0.001	
			(0.001)	
Social connectedness · GDP growth decile 4			0.001	
			(0.001)	
Social connectedness · GDP growth decile 6			0.000	
			(0.001)	
Social connectedness · GDP growth decile 7			0.000	
			(0.001)	
Social connectedness · GDP growth decile 8			0.001	
			(0.001)	
Social connectedness · GDP growth decile 9			0.001	
			(0.001)	
Social connectedness · GDP growth decile 10			0.003***	
			(0.001)	
Social connectedness · GDP growth differential decile 2				-0.000
				(0.001)
Social connectedness · GDP growth differential decile 3				-0.000
				(0.001)
Social connectedness · GDP growth differential decile 4				-0.000
				(0.001)
Social connectedness · GDP growth differential decile 5				0.000
				(0.001)
Social connectedness · GDP growth differential decile 6				0.001
				(0.001)
Social connectedness · GDP growth differential decile 7				0.003***
				(0.001)
Social connectedness · GDP growth differential decile 8				0.003***
				(0.001)
Social connectedness · GDP growth differential decile 9				0.002*
				(0.001)
Social connectedness · GDP growth differential decile 10				0.002**
				(0.001)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
Physical and cultural distances	Yes	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684	66,684
Adj. R^2	0.551	0.525	0.525	0.525
Adj. R^2 within	0.181	0.132	0.132	0.132

Table 8: Information sensitivity of loans across loan types

Based on the regressions reported in this table, we analyze how the relationship between bank lending and social connectedness depends on banks' screening incentives. Columns 1, 2, 4, and 5 use sample splits based on government guarantees and securitization. The number of observations is identical to the full sample as we perform the sample splits at the loan level, meaning before aggregating the loan volumes at the county-pair level. Columns 3 and 6 use interaction terms with continuous variables instead of sample splits. "Guaranteed share" refers to the share of the loan volume subject to government guarantees and is also included as a single term in column 3. "Sold share" refers to the share of the loan volume sold off book and is also included as a single term in column 6. Control variables are the physical and cultural distances, 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. Table A1 defines the variables. 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 mortgage 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	0.012***	0.016***	0.008**	0.022***
	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)
Social connectedness · Guaranteed share			-0.032***			
			(0.004)			
Social connectedness · Sold share						-0.038***
						(0.003)
Source county FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes	Yes	Yes
Physical and cultural distances	Yes	Yes	Yes	Yes	Yes	Yes
County-pair-level control variables	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	0.295	0.337	0.363	0.317	0.583
Adj. R^2 within	0.065	0.041	0.255	0.102	0.044	0.530

Table 9: Ex-ante lending risk and ex-post loan performance

This table reports loan-level regressions of borrower characteristics, loan characteristics, and loan performance on social connectedness, which we use to explore how social connectedness relates to the ex-ante and ex-post riskiness of loans. "Combined LTV" is the combined loan-to-value ratio of all mortgages on the borrower's property. "Delinquent" is a dummy variable which indicates mortgages that have been at least 90 days past due at least once. For readability of the coefficients, the delinquent dummy enters regressions as a share of its standard deviation. All columns control for the physical and cultural distances between borrower and lender. The additional loan characteristics in columns 3 to 5 are the natural logarithm of the loan volume, the debt-to-income ratio, and a binary variable indicating first-time home buyers. Table A1 defines the variables. 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)
Timing relative to lending decision:	Ex ante		During	Ex post	
Dependent variable:	FICO score	Combined LTV	Interest rate	Delinquent	log(Unpaid balance)
Social connectedness	0.004	-0.017	-0.068**	0.001	-0.020*
	(0.036)	(0.012)	(0.031)	(0.001)	(0.011)
Social connectedness · Not in default					0.017
					(0.011)
FICO score			-0.092***	-0.004***	0.004***
			(0.008)	(0.000)	(0.001)
Combined LTV			0.454***	0.003***	-0.005*
			(0.024)	(0.001)	(0.003)
Interest rate				0.001***	0.002
				(0.000)	(0.002)
Delinquent					0.134***
					(0.041)
Bank FE	Yes	Yes	Yes	Yes	Yes
Origination year FE	Yes	Yes	Yes	Yes	Yes
Physical and cultural distances	Yes	Yes	Yes	Yes	Yes
Additional loan characteristics	No	No	Yes	Yes	Yes
No. of obs.	20,760	20,760	20,760	20,760	20,760
Adj. R^2	0.024	0.039	0.770	0.092	0.083
Adj. R^2 within	0.000	0.000	0.085	0.073	0.005

Table 10: Real effects of social proximity to bank capital

This table reports county-level analyses of the real effects of social proximity to bank capital (see Equation 4). "Loan volume" is the volume of loans to small and medium-sized enterprises. Control variables account for the physical and cultural proximity to bank capital, industry shares, and, in columns 4 and 5, commuting and migration. Table A1 defines the variables. The parentheses contain standard errors clustered at the county level. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	log(Loan volume)	Real GDP growth		log(Employment)	
Social proximity	0.047**	0.850***	0.638**	0.004**	0.003
	(0.021)	(0.274)	(0.263)	(0.002)	(0.002)
Social proximity · Small-firm percentage			0.0373**		0.0003**
			(0.0159)		(0.0001)
Small-firm percentage			0.919***		0.000
			(0.302)		(0.001)
County FE	Yes	Yes	Yes	Yes	Yes
State-time FE	Yes	Yes	Yes	Yes	Yes
Physical and cultural proximity	Yes	Yes	Yes	Yes	Yes
Additional control variables	Yes	Yes	Yes	Yes	Yes
No. of obs.	24,152	24,161	24,161	24,161	24,161
Adj. R^2	0.968	0.240	0.241	0.999	0.999
Adj. R^2 within	0.004	0.131	0.132	0.136	0.137

Table 11: Social connectedness and the allocation of bank lending: heterogeneities across distances

The regressions reported in this table estimate the dependence of the relationship between bank lending and social connectedness on distances and connectedness. Column 1 restates the baseline regression from Table 4, 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. 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. Table A1 defines the variables. 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**	0.022***	0.002*
	(0.001)	(0.003)	(0.004)	(0.001)
Social connectedness^2		-0.000		
		(0.000)		
Social connectedness $\cdot$ Physical distance			-0.004***	
			(0.001)	
Social connectedness $\cdot$ Cultural distance				0.0004***
				(0.0001)
Physical distance	-0.267***	-0.264***	-0.251***	-0.247***
	(0.051)	(0.055)	(0.051)	(0.053)
Cultural distance	-0.016**	-0.015**	-0.017**	-0.021***
	(0.007)	(0.007)	(0.007)	(0.008)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
County-pair level control variables	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	0.133	0.134

Table 12: Alternative measures of physical distance

This table illustrates the robustness of our results with respect to alternative measures of physical distance. Column 1 restates the baseline regression from Table 4, column 4. Columns 2 to 4 re-estimate the baseline regressions with the 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. Table A1 defines the variables. 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***
	(0.001)	(0.001)	(0.001)	(0.001)
Physical distance	-0.267***			
	(0.051)			
Physical distance (population-weighted centroids)		-0.253***		
		(0.051)		
Highway travel costs			-0.260***	
			(0.057)	
All modes travel costs				-0.263***
				(0.057)
Cultural distance	-0.016**	-0.016**	-0.017**	-0.018**
	(0.007)	(0.007)	(0.007)	(0.007)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
County-pair-level control variables	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	0.131	0.130	0.130

Table 13: Alternative specifications of cultural distance

This table illustrates the robustness of our results with respect to measures of cultural distance. Column 1 restates the baseline regression from Table 4, column 4. Columns 2 to 4 re-estimate the baseline regressions with alternative specifications of cultural distance. "Individual culture controls" refers to the 39 variables used to construct our measure of cultural distance (see Table A2). Column 5 uses the absolute difference of the vote share for the Republican candidate in the 2016 presidential election to proxy for culture. Column 6 adds a measure of the average cultural heterogeneity within the source and destination counties and its interaction with social connectedness to our baseline specification. 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. Table A1 defines the variables. 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)	(6)
Dependent variable:	log(Volume of SME loans)					
Social connectedness	0.007***	0.007***	0.007***	0.007***	0.007***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Social connectedness · Cultural het.						-0.009
						(0.007)
Physical distance	-0.267***	-0.285***	-0.277***	-0.314***	-0.295***	-0.273***
	(0.051)	(0.052)	(0.053)	(0.040)	(0.047)	(0.048)
Cultural distance	-0.016**					-0.015**
	(0.007)					(0.007)
Cultural distance: ethnic ancestry		-0.008	-0.027			
		(0.043)	(0.042)			
Cultural distance: racial origin		-0.014	-0.087*			
		(0.054)	(0.045)			
Cultural distance: religious beliefs		-0.003	-0.003			
		(0.059)	(0.060)			
Cultural distance: social environment		-0.125***				
		(0.048)				
Vote-share differential					-0.697**	
					(0.305)	
Source county FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual culture controls	No	No	No	Yes	No	No
No. of obs.	66,684	66,684	66,684	66,684	66,377	66,684
Adj. R^2	0.525	0.525	0.525	0.531	0.525	0.525
Adj. R^2 within	0.132	0.133	0.131	0.144	0.133	0.132

Appendix: Additional Tables

Table A1: Variable definitions and data sources

This table provides variable definitions and data sources. For descriptive statistics on the main sample of cross-county loans to small and medium-sized enterprises see Table 1. Tables 2, 3, and OA3 report descriptive statistics of the additional data used to analyze heterogeneities across the information sensitivity of loans in Section 4.1.2, the riskiness of loans in Section 4.2, and real effects in Section 4.3.

Variable name	Description
Connectedness and distance measures	
Social connectedness	Relative probability of friendship links between source and destination county; scaled to [0;100]; source: Bailey, Cao, Kuchler, and Stroebel (2018).
Physical 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]; source: own calculation as described in Section 3.1.
Bank-lending data at county-pair level	
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.
Loans without guarantees	Volume of newly originated mortgage loans that are not subject to government guarantees; enters regressions in logs; source: HMDA.
Loans with guarantees	Volume of newly originated mortgage loans that are subject to government guarantees; enters regressions in logs; source: HMDA.
Guaranteed share	Share of the mortgage loan volume subject to government guarantees; source: HMDA.
Loans kept on book	Volume of newly originated mortgage loans that are kept on the originating bank's balance sheet; enters regressions in logs; source: HMDA.
Loans sold (off book)	Volume of newly originated mortgage loans that are sold and thus not kept on the bank's balance sheet; enters regressions in logs; source: HMDA.
Sold share	Share of the mortgage loan volume sold off the books; source: HMDA.
Main control variables at county-pair level	
Common border	Indicator variable equal to 1 if source and destination county are direct neighbors; source: U.S. Census Bureau.
GDP growth differential	Absolute value of the county-pair difference in the average real GDP growth during the last three years; in percentage points; source: U.S. Bureau of Economic Analysis.
Gross commuting	Share of the county-pair population commuting from source to destination county or vice versa; in %; source: U.S. Census Bureau.
Gross migration	Share of the county-pair population which migrated from source to destination county or vice versa; in %; U.S. Census Bureau.
Gross trade	Gross value of trade between source and destination county; in million USD; source: U.S. Census Bureau.
Unemployment differential	Absolute value of the county-pair difference in the unemployment rate; in percentage points; source: U.S. Bureau of Labor Statistics.
Same state	Binary variable equal to 1 if source and destination county are located in the same state; source: NBER's county distance database.

(*table continued on next page*)

Table A1 - *continued*

Variable name	Description
Instrumental variables	
Historical travel costs	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 4.1.1 and Equation 6); source: own calculation and U.S. Census Bureau.
Same highway	Binary variable equal to 1 if source and destination county are connected by the same highway; source: Baum-Snow (2007).
Years since highway construction	Number of years for which source and destination county have been connected by the same highway; source: Baum-Snow (2007).
Variables at county-pair level used in interaction terms	
Bank size	Natural logarithm of the loan-volume-weighted average balance sheet size of banks that lend from source to destination county; source of total assets data: FDIC.
Borrower's volatility exposure	Destination county's exposure to industry-level output volatility calculated using county's industry shares to weight the U.S.-wide output volatility of industries; source: U.S. Bureau of Economic Analysis.
GDP growth	Destination county's average real GDP growth during the last three years; in %; source: U.S. Bureau of Economic Analysis.
Further variables at county-pair level	
Culture: ethnic ancestry	Index quantifying a county pair's cultural distance based on ethnic ancestries; calculation described in Section 3.1; source: own calculation.
Culture: racial origin	Index quantifying a county pair's cultural distance based on racial origins; calculation described in Section 3.1; source: own calculation.
Culture: religious beliefs	Index quantifying a county pair's cultural distance based on religious beliefs; calculation described in Section 3.1; source: own calculation.
Culture: social environment	Index quantifying a county pair's cultural distance based on social environments; calculation described in Section 3.1; source: own calculation.
Cultural heterogeneity	Index quantifying the sum of the "cultural distance" between census tracts in the source county and the "cultural distance" between census tracts in the destination county; source: own calculation.
All modes travel costs	Costs of traveling from source to destination county via the cheapest combination of highways, roads, and waterways; source: Oak Ridge National Laboratory's National Transportation Center.
Highway travel costs	Costs of traveling from source to destination county via highways; source: Oak Ridge National Laboratory's National Transportation Center.
Physical distance (pop.-weighted centroids)	Great-circle distance in miles based on *population-weighted* county centroids; source: U.S. Census Bureau, 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.
Inverse population product	One over the product of the source and destination county populations; source: U.S. Census Bureau.

(*table continued on next page*)

Table A1 - *continued*

Additional loan-level data	
Combined LTV	Combined loan-to-value ratio of all mortgages on the borrower's property; in %; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Delinquent	Indicator variable equal to 1 if the loan is at least 90 days past due; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
DTI	Borrower's debt-to-income ratio; in %; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
FICO score	Borrower's credit score; higher values signal higher creditworthiness; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Interest rate	Loan's interest rate at the time of origination; in basis points; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
First-time buyer	Indicator variable equal to 1 if the borrower is buying a home for the first time; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Not in default	Indicator variable equal to 1 if the loan is not in default and not at least 2.5 years past due; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Original loan amount	Loan amount at the time of origination; enters regressions in logs; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Unpaid balance	Current amount still owed by the borrower; enters regressions in logs; source: Fannie Mae's and Freddie Mac's Single Family Loan-Level Data Sets.
Additional county-level data for the real-effects analysis	
Social proximity to bank capital	County i's social proximity to bank capital at time t is defined as $\sum_j Social\ connectedness_{i,j} \cdot Total\ assets_{j,t}$; total assets is the sum of total assets of all banks with headquarters in county j; scaled to a standard deviation of 1; sources: Bailey, Cao, Kuchler, and Stroebel (2018), FDIC, own calculation.
Physical proximity to bank capital	County i's physical proximity to bank capital at time t is defined as $\sum_j (Physical\ distance)^{-1}_{i,j} \cdot Total\ assets_{j,t}$; total assets is the sum of total assets of all banks with headquarters in county j; scaled to a standard deviation of 1; sources: NBER's county distance database, FDIC, own calculation.
Cultural proximity to bank capital	County i's cultural proximity to bank capital at time t is defined as $\sum_j (Cultural\ distance)^{-1}_{i,j} \cdot Total\ assets_{j,t}$; total assets is the sum of total assets of all banks with headquarters in county j; scaled to a standard deviation of 1; source: own calculation.
Employment	Number of employed people; enters regressions in logs; source: U.S. Bureau of Labor Statistics.
Real GDP growth	County-level real GDP growth; in %; source: U.S. Bureau of Economic Analyses.
Small-firm percentage	Share of small firms (=less than 20 employees) in the destination county; calculated as the nationwide share of small firms per industry weighted with a county's industry shares; in %; source: U.S. Census Bureau's Statistics of U.S. Businesses.

Table A2: Measurement of cultural distance: variables, categories, and subcategories

The table lists the variables for the measurement of the cultural distance between counties and associates them with Lieske's (1993) four dimensions of regional subcultures: ethnic ancestry, racial origin, religious beliefs, and the social environment. This environment is further divided into subcategories for weighting purposes. Section 3.1 explains the construction of our cultural distance measure in detail.

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
% Sub-Saharan African			% females in labor force	**Racial diversity**
			Income inequality	Gini coefficient of
			Gini coefficient	racial origins

The Role of Social Networks in Bank Lending

Online Appendix

Oliver Rehbein†

University of Bonn

Simon Rother‡

University of Bonn

October 1, 2020

(Please click here for the latest version)

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

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

Figure OA1: Marginal effect of social connectedness across time

The figure displays the evolution of the marginal effect of social connectedness on log loan volumes over time. In line with the specification in Table 5, column 2, the results are obtained while controlling for source-county-time fixed effects, destination-county-time fixed effects, physical and cultural distances, and the additional macroeconomic control variables. However, unlike in that regression, social connectedness is interacted with indicator variables for each year.

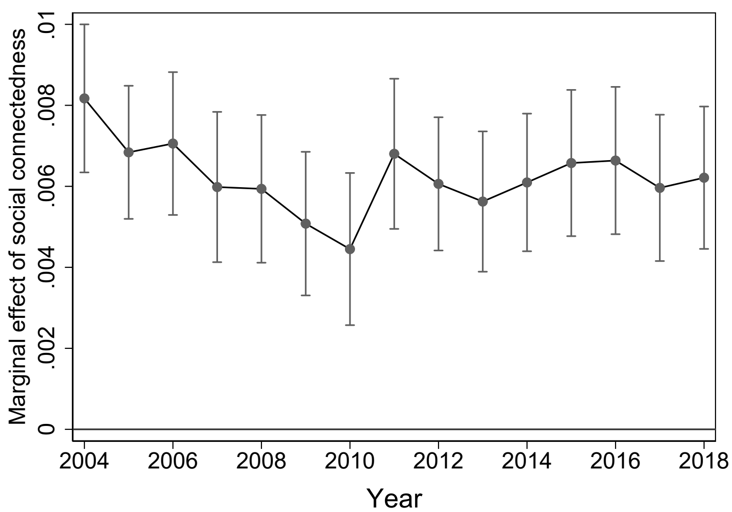

Figure OA2: Marginal effect of social connectedness across the information sensitivity of loans

The figures display the marginal effect of social connectedness on log loan volumes in dependence of the variables indicated below each figure. Figures (a) and (b) are based on the estimates reported in Table 7, columns 1 and 2. Figures (c) and (d) are based on the estimates reported in Table 8, columns 3 and 6. Table A1 defines the variables. For summary statistics, see Tables 1 and OA3.

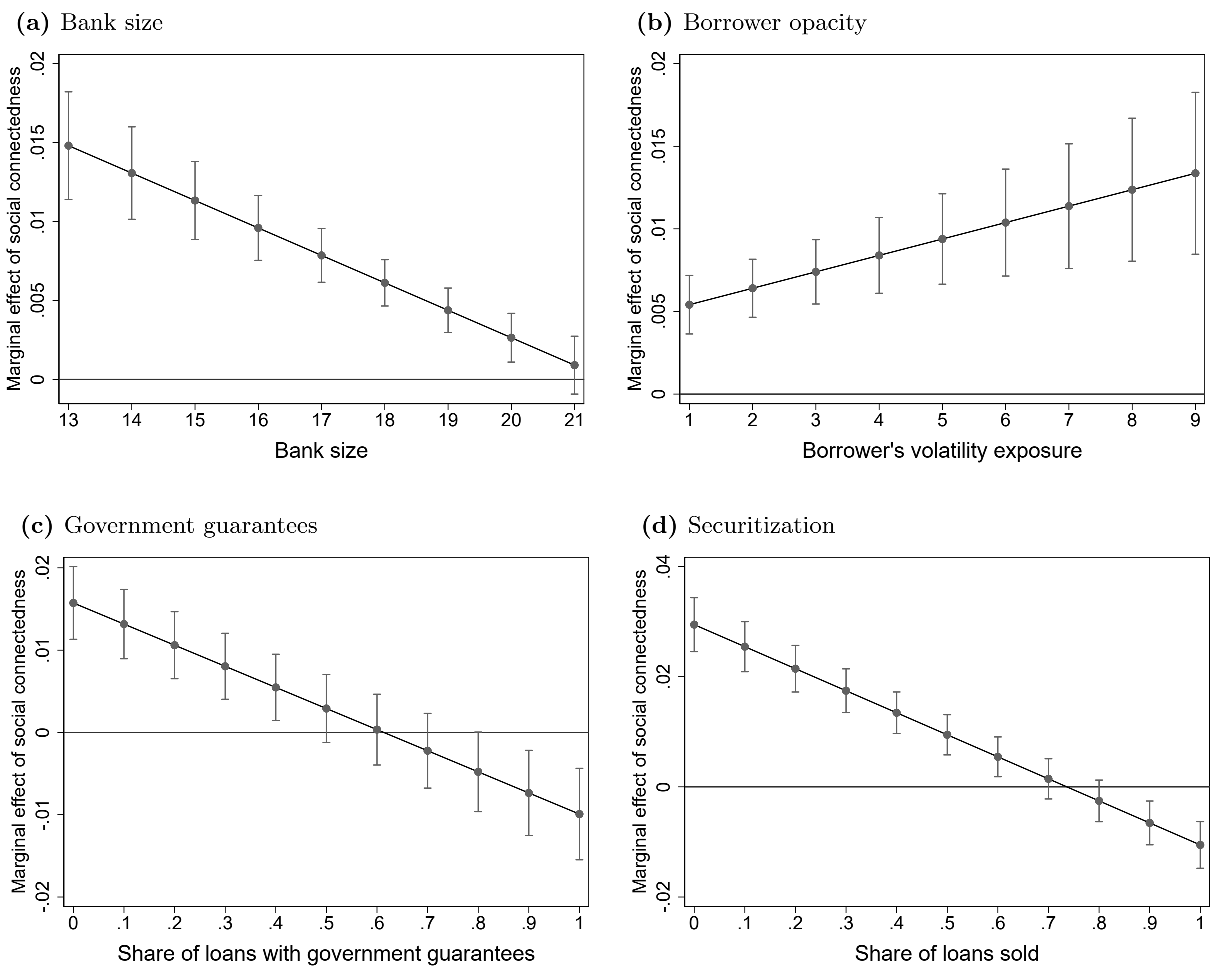

Table OA1: The role of distance and connectedness: alternative clustering

The regressions in this table assess the robustness of our results with respect to alternative approaches to the clustering of standard errors. Column 1 restates our baseline regression reported in column 4 of Table 4, 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. Table A1 defines the variables. The parentheses contain the 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	-0.269***	-0.269***	-0.269***
	(0.051)	(0.048)	(0.051)
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
County-pair-level control variables	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684
Adj. R^2	0.525	0.525	0.525
Adj. R^2 within	0.132	0.132	0.132

Table OA2: Social connectedness and the allocation of bank lending: headquarter locations

This table illustrates the robustness of our results with respect to the definition of the source county of loans. Column 1 restates the baseline regression from Table 4, column 4, where source counties are based on banks' branch networks. In column 2, we re-estimate our baseline regression but determine source counties based on banks' 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. Table A1 defines the variables. 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***
	(0.001)	(0.002)
Physical distance	-0.267***	-0.381***
	(0.051)	(0.087)
Cultural distance	-0.016**	-0.027***
	(0.007)	(0.009)
Source county FE	Yes	Yes
Destination county FE	Yes	Yes
County-pair-level control variables	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	-0.19	-0.17
Cultural distance	-0.09	-0.06

Table OA3: Descriptive statistics: mortgage-lending sample

The table displays descriptive statistics for the county-pair-level mortgage loan sample used to analyze heterogeneities across the information sensitivity of loans in Table 8. All variables are at the county-pair level. For instance, "volume of mortgage loans" equals the volume of loans provided by all bank branches in a source county to all borrowers in the destination county. Table A1 provides variable definitions and data sources.

	N	Mean	Median	SD	Min	Max
Connectedness and distances						
Social connectedness	34,483	33	4	42	0	100
Cultural distance	34,483	14	13	7	0	47
Physical distance [miles]	34,483	452	272	501	5	4,898
log(Physical distance)	34,483	5.4	5.6	1.4	1.6	8.5
Lending data						
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
Loans without guarantees [log(Volume of)]	34,483	8.8	12.0	6.1	0.0	19.8
Loans with guarantees [log(Volume of)]	34,483	4.2	0.0	6.1	0.0	18.7
Guaranteed share	34,483	0.2	0.0	0.4	0.0	1.0
Loans kept on book [log(Volume of)]	34,483	4.9	0.0	6.2	0.0	19.3
Loans sold [log(Volume of)]	34,483	7.8	11.9	6.5	0.0	19.7
Sold share	34,483	0.5	0.7	0.5	0.0	1.0
County-pair-level control variables						
Common border	34,483	0.2	0.0	0.4	0.0	1.0
GDP growth differential [ppt.]	34,483	3.2	2.4	3.2	0.0	37.1
Gross trade [million USD]	34,483	68	21	111	0	1,056
Gross migration [%]	34,483	0.06	0.00	0.17	0.00	3.35
Gross commuting [%]	34,483	0.3	0.0	0.9	0.0	16.9
Same state	34,483	0.3	0.0	0.5	0.0	1.0
Unemployment differential [ppt.]	34,483	1.3	1.0	1.2	0.0	20.3

Table OA4: Social connectedness and the allocation of bank lending: coefficients of control variables

The table restates the regressions from Table 4 while additionally reporting the coefficients of control variables. The dependent variable is the logarithm of the volume of cross-county loans to small and medium-sized enterprises. Table A1 defines the variables. 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	0.012***			0.007***
	(0.001)			(0.001)
Physical distance		-0.389***		-0.267***
		(0.041)		(0.051)
Cultural distance			-0.034***	-0.016**
			(0.007)	(0.007)
Same state	0.339***	0.296***	0.777***	0.107
	(0.070)	(0.084)	(0.072)	(0.069)
Common border	0.774***	0.804***	1.078***	0.655***
	(0.050)	(0.076)	(0.050)	(0.061)
GDP growth differential	-0.008	-0.007	-0.006	-0.006
	(0.007)	(0.008)	(0.008)	(0.008)
Unemployment differential	-0.015	-0.019	-0.007	-0.002
	(0.022)	(0.022)	(0.021)	(0.020)
Gross trade	0.001***	0.000	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gross commuting	0.089***	0.074***	0.110***	0.082***
	(0.027)	(0.026)	(0.026)	(0.025)
Gross migration	0.252*	0.302**	0.241*	0.204
	(0.132)	(0.129)	(0.134)	(0.135)
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	0.519	0.522	0.515	0.525
Adj. R^2 within	0.121	0.126	0.114	0.132

Table OA5: Social connectedness and the allocation of bank lending within and across states

This table analyzes the relationship between bank lending and social connectedness within and across states. Column 1 restates our baseline regression from Table 4, column 4. Columns 2 and 3 introduce interactions of social connectedness and the distance measures with the same-state indicator 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. Table A1 defines the variables. 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)		
Social connectedness	0.007***	0.005***	0.006***
	(0.001)	(0.001)	(0.001)
Social connectedness · Same state		0.003***	0.000
		(0.001)	(0.002)
Physical distance	-0.267***	-0.284***	-0.241***
	(0.051)	(0.053)	(0.057)
Physical distance · Same state			-0.300***
			(0.068)
Cultural distance	-0.016**	-0.016**	-0.021***
	(0.007)	(0.007)	(0.008)
Cultural distance · Same state			0.031***
			(0.008)
Source county FE	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes
No. of obs.	66,684	66,684	66,684
Adj. R^2	0.525	0.525	0.526
Adj. R^2 within	0.132	0.132	0.135

Table OA6: Social connectedness and the probability of bank lending at shorter distances

Columns 1 and 3 restate the regressions from Table 5, columns 4 and 5. Columns 2 and 4 re-estimate the same specification but restrict the sample to county pairs that are closer to each other than the median distance between counties in the United States, 776 miles, as banks rarely lend across longer distances. 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. Table A1 defines the variables. 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)		Lending indicator	
Maximum distance between county pairs:	Unrestricted	776 miles	Unrestricted	776 miles
Social connectedness	0.0033***	0.0029***	0.0002***	0.0002***
	(0.000)	(0.000)	(0.000)	(0.000)
Physical distance	-0.1432***	-0.2373***	-0.0078***	-0.0128***
	(0.013)	(0.018)	(0.001)	(0.001)
Cultural distance	-0.0042***	-0.0062***	-0.0002**	-0.0003***
	(0.001)	(0.002)	(0.000)	(0.000)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes	Yes
No. of obs.	9,526,482	4,763,196	9,526,482	4,763,196
Adj. R^2	0.337	0.299	0.332	0.297
Adj. R^2 within	0.077	0.092	0.068	0.082

Table OA7: Instrumental variable approaches: first-stage regressions

This table reports the first-stage regressions of the instrumental-variable estimates reported in Table 6. Table A1 defines the variables. 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	3.783***			
	(0.518)			
Years since highway construction		0.071***		
		(0.010)		
Historical travel costs			4.328***	
			(0.377)	
Relative Facebook county rank				163.117***
				(13.265)
Physical distance	-12.338***	-12.291***	-13.487***	-15.344***
	(1.580)	(1.579)	(1.466)	(0.939)
Cultural distance	-0.434***	-0.433***	-0.378***	-0.769***
	(0.053)	(0.053)	(0.055)	(0.073)
Same state	27.519***	27.521***	22.623***	26.351***
	(1.843)	(1.843)	(1.463)	(1.691)
Common border	20.717***	20.726***	13.235***	13.532***
	(1.288)	(1.288)	(1.302)	(1.346)
GDP growth differential	0.009	0.009	-0.064	0.009
	(0.065)	(0.065)	(0.048)	(0.086)
Unemployment differential	-0.734***	-0.733***	-0.635***	-0.698***
	(0.213)	(0.213)	(0.173)	(0.218)
Gross trade	-0.004	-0.004	-0.004	0.011***
	(0.003)	(0.003)	(0.003)	(0.004)
Gross commuting	0.019	0.036	-1.443***	-7.314***
	(0.377)	(0.377)	(0.375)	(0.911)
Gross migration	5.565***	5.575***	5.281***	5.331
	(1.935)	(1.938)	(1.995)	(6.011)
Present-day travel costs	1.285	1.226	-13.236***	
	(1.715)	(1.712)	(1.791)	
Population control				-0.000***
				(0.000)
Source county FE	Yes	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes	Yes
No. of obs.	66,647	66,647	56,223	56,852
Adj. R^2	0.858	0.858	0.882	0.814
Adj. R^2 within	0.674	0.674	0.734	0.535

Table OA8: Instrumental variable approaches: alternative specifications

This table provides robustness checks for our instrumental variable approaches. Columns 1 and 2 re-estimate the instrumental-variable regressions reported in columns 1 and 2 of Table 6 while excluding counties in which the share of the population that lives in urban areas exceeds the 75^{th} percentile of its distribution. 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 6, 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. Table A1 defines the variables. 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	0.034***	0.037***	0.128*
	(0.013)	(0.014)	(0.076)
Physical distance	-0.252	-0.223	0.861
	(0.163)	(0.170)	(0.687)
Cultural distance	-0.002	-0.001	0.045
	(0.009)	(0.009)	(0.035)
Source county FE	Yes	Yes	Yes
Destination county FE	Yes	Yes	Yes
County-pair-level control variables	Yes	Yes	Yes
Present-day travel costs	Yes	Yes	No
No. of obs.	49,725	49,725	56,852
Adj. R^2	0.485	0.478	-0.010
Adj. R^2 within	0.083	0.069	-0.944
Instrument (1st stage)	3.427***	0.064***	16.778***
	(0.000)	(0.000)	(0.001)
F-value (1st stage)	41.241	37.386	10.6

Table OA9: Social connectedness and the allocation of bank lending: SME loans vs. mortgage loans

This table compares our baseline estimates obtained based on the sample of cross-county loans to small and medium-sized enterprises with the corresponding estimates based on the mortgage-loan sample. Column 1 restates our baseline regression reported in column 4 of Table 4. Column 2 reports the results for the mortgage loan sample. The standardized beta coefficients at the end of the table allow a meaningful comparison of the size of the estimates as they express the effect of a 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. Table A1 defines the variables. 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)
Dep. var.: log(Volume of ...)	SME loans	Mortgage loans
Social connectedness	0.007***	0.009***
	(0.001)	(0.003)
Physical distance	-0.267***	-0.462***
	(0.051)	(0.109)
Cultural distance	-0.016**	-0.037**
	(0.007)	(0.015)
Source county FE	Yes	Yes
Destination county FE	Yes	Yes
County-pair-level control variables	Yes	Yes
No. of obs.	66,684	34,483
Adj. R^2	0.525	0.157
Adj. R^2 within	0.132	0.051
Standardized beta coefficients		
Social connectedness	0.12	0.07
Physical distance	-0.17	-0.12
Cultural distance	-0.06	-0.05