Social Interactions and Peer-to-Peer Lending Decisions*

Linda Allen†, Lin Peng‡, Yu Shan§

June 1, 2019

Abstract

We examine the effects of social connectivity on the demand for and supply of consumer and small business loans on peer-to-peer (P2P) FinTech sites such as LendingClub. P2P loan demand increases when geographically distant, but socially connected areas have large amounts of past P2P borrowing activity. Both approval rates and quality (as measured by loan grade and interest rates) are higher the greater an area's aggregate online social connections. Performance (i.e., reductions in defaults or delayed payments) is enhanced by social connectivity indicating that information diffusion through online social networks improves lending outcomes for both high and low risk borrowers.

JEL classification: G20, G21, G24, G28.

Keywords: Social Interaction, P2P Lending, Information Asymmetry.

---

*We thank Vikas Agarwal, Gennaro Bernile, Ling Cen, Youngmin Choi, Tarun Chordia, Phil Dvbvig, Sonali Hazarika, Armen Hovakimian, Maggie Hu, Jiasun Li, Xiuming Martin Roni Michaely, Vesa Pursiainen, Yao Shen, Yang Shi, Johannes Stroebel Bohui Zhang, Dexin Zhou, and other conference participants at the 2018 Conference on Fintech, Social Finance, and Financial Stability, 2019 Hong Kong University FinTech conference, the 2nd Sun Yat-Sen University Finance International Conference and Baruch Seminar for helpful comments and suggestions. We are responsible for remaining errors.

†William F. Aldinger Chair in Banking and Finance, Bert W. Wasserman Department of Economics and Finance, Zicklin School of Business, Baruch College, CUNY, One Bernard Baruch Way, New York, NY 10010. E-mail: linda.allen@baruch.cuny.edu.

‡Krell Chair of Finance, Professor of Economics and Finance, Bert W. Wasserman Department of Economics and Finance, Zicklin School of Business, Baruch College, CUNY, One Bernard Baruch Way, New York, NY 10010. E-mail: lin.peng@baruch.cuny.edu.

§Ph.D. candidate in finance, Bert W. Wasserman Department of Economics and Finance, Zicklin School of Business, Baruch College, CUNY, One Bernard Baruch Way, New York, NY 10010. E-mail: yu.shan@baruch.cuny.edu.

A fundamental observation about human society is that people who communicate regularly with one another think similarly.... Word-of-mouth transmission of ideas appears to be an important contributor to day-to-day or hour-to-hour stock market fluctuations.

- Robert Shiller, Irrational Exuberance

1. Introduction

While classical investment theories assume that investment ideas are transmitted among investors through asset prices and quantities in impersonal markets, recent theoretical models and empirical evidence suggests that various forms of direct social interaction also impact investment decisions. For example, individual investors often purchase stocks upon recommendation of their friends and relatives (Shiller and Pound, 1989), and mutual fund managers are more likely to buy a particular stock if other managers in the same city are buying that same stock (Hong, Kubik, and Stein, 2005).[1] In the past two decades, the proliferation of online social media and social networking service companies such as Facebook and Linkedin provide innovative and efficient tools for social interaction, thus potentially impacting investment and lending decisions. However, the empirical case for the importance of these social media sites on economic decision making has not yet been well developed.

In this paper, we explore the economic role of social interactions by investigating how they affect individual borrowing and lending decisions on the largest peer-to-peer (P2P) lending platform in the U.S., LendingClub. Our analyses focus on both the demand side and the supply side: how borrowers' social interaction intensity affects their borrowing decisions and how it affects loan approval, pricing, and performance. We study these questions using a novel measure of social interactions, "Social Connectedness Index" (SCI), which is produced by Bailey, et al. (2017) using the aggregated and anonymized information from the universe

[1] For other studies showing how social interactions impact investment decisions, see Kelly and O'Grada (2000), Duflo and Saez (2003), Hong, Kubik, and Stein (2004), Massa and Simonov (2005), Ivković and Weisbenner (2007), Cohen, Frazzini, and Malloy (2008, 2010), and Shive (2010).

of friendship links between all Facebook users.[2] We show that social interactions not only enable borrowers to learn from friends' past borrowing experiences and make their own borrowing decisions, but also allow lenders to extract valuable information about the borrowers and make improved investment decisions. Our findings supports the argument that social interactions affect individual economic decisions through the channel of enhancing information transmission and diffusion.

The Facebook SCI proposed by Bailey et al. (2017) provides a snapshot of the county-pair intensities of social interactions for all counties in the U.S.. Facebook is the world's largest online social network with more than 2.2 billion monthly active users worldwide and 169.5 million monthly active users in the U.S. as of 2018. In January 2018, around two-thirds (68%) of U.S. adults use Facebook and 70% of users visit the site on a daily basis,[3] making Facebook the most popular social media site in the U.S.. People primarily use Facebook to connect, share, discover, and communicate with their real-world friends and acquaintances. Its immense scale, great market penetration, and high popularity make Facebook's friendship linkage data a reasonable representation of real-word friendship interactions.

We first hypothesize that borrowers can learn from their socially connected friends about their past borrowing experience on LendingClub and make borrowing decisions accordingly. It essentially argues that social interactions facilitate the spread of new financial products across geographic areas by disseminating the customers' perception and recognition of these products through social network. Specifically, we examine how the number of loan applications in an area changes when their socially connected areas experienced high growth in loan application, origination, or high approval rates. In order to confirm a causal relationship between friend areas' past borrowing experience and own borrowing decisions, we need to rule out a confounding effect which arises when borrowers make borrowing decisions by extrapolating from their own borrowing experiences, which could be correlated with the bor-

[2] We acknowledge Michael Bailey, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel for providing the county-level Facebook friendship link data.

[3] "Social Media Use in 2018", http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

rowing experiences of friend areas. To extract only variation in friend areas' past borrowing experiences that are orthogonal to an area's past own experience, we instrument for the past experience of all friend areas with the experiences of only geographically distant friend areas (out-of-state areas or out-of-commuting-zone areas).

We find that the past LendingClub borrowing experiences of friend areas have a quantitatively large impact on the area's subsequent borrowing decisions. An area experiences higher growth in loan applications when their distant friend areas experienced high growth in loan application. On average, if the growth of applications per one thousand people increases by 10 applications between 2010 to 2013 in friend areas, then the growth of applications per one thousand people increases by 2.8 applications between 2013 to 2016. The results are robust if we replace friend areas' past loan application experiences with loan origination, average approval rates, or increases in approval rates. The results imply that local borrowers are likely to form their own perceptions about P2P financing by learning their friends past borrowing experiences and then make their own borrowing decisions accordingly.

We then hypothesize that social network interaction may also impact lending outcomes. Social interactions can affect lending outcomes through two competing channels: Emotional and Cognitive Biases Channel versus Information Dissemination Channel. The emotional bias channel hypothesizes that friendship connections create emotional attachments, which makes a person feel concern, caring, or affection for the other person. This attachment may hinder the investor's judgment and interfere with the ability to make rational investment decisions. That is, online lenders may lend to borrowers in the same geographic areas as their Facebook friends even when these socially connected borrowers are not particularly creditworthy. The cognitive bias channel suggests that strong social connections may generate persuasion bias (DeMarzo, Vayanos, and Zwiebel, 2003) and overconfidence bias (De Bondt and Thaler, 1995; Barber and Odean, 2000) that may lead to poor investment decisions. DeMarzo et al. (2003) show that the failure of receivers to account for possible repetition in the messages they hear from others (i.e., persuasion bias) plays an important role in the process of social opinion formation. Thus, one's influence on group opinions depends not only on accuracy, but also on popularity and the extent of one's social connections. In our context, loan requests from a well-connected area may be more easily received and accepted by investors simply because of familiarity generated through online interactions with others in the borrower's geographic area. Overconfidence bias may arise when well-connected investors erroneously believe that they have an edge that others do not have. DeMarzo et al. (2003) and Barber and Odean (2000) show that excessive trading is often attributed to investor overconfidence. Online social interactions may exacerbate this trading aggressiveness (Han, Hirshleifer, and Walden, 2018). In our context, lenders may mistake emotional and cognitive biases for information, and therefore irrationally favor borrowers from socially connected regions.

In contrast, the information dissemination channel hypothesizes that tighter social connections across regions may facilitate communication, reduce uncertainty, and reduce cultural differences, thereby lowering the cost of information dissemination and facilitating learning (Golub and Jackson, 2012). Further, literature on informal lending markets (such as microfinance in rural areas) suggests that reputational concerns within localized areas enhance credit outcomes (Ghosh and Ray, 2016). Social media can act as a reputation enforcement mechanism if repayment performance improves because of peer pressure through publicly disclosed defaults. For example, Ge, Gu, and Feng (2017) find that loan performance is better for a group of borrowers that choose to disclose their social media identities. Thus, stronger online social networks may help overcome information asymmetries between borrowers and lenders, and enable investors to more accurately screen loan requests. For example, online social connections enable the rapid and accurate dissemination of information about the economic environment in geographically dispersed regions. This could allow investors to better assess the creditworthiness of loan applications from these regions. Similar to physical proximity promotion of interactions among mutual fund managers in Coval and Moskowitz (1999), we hypothesize that online proximity facilitates information transmission, thereby reducing both information acquisition costs and the perceived risk of loan applications from socially connected areas.

We test both competing hypotheses by first establish an association between loan approval and social connectivity. We examining both loan-level approval and area-level approval rates while controlling for borrower and loan characteristics as well as local social, economic, and demographic characteristics. We find that loan applications from more socially connected areas are more likely to get approval, obtain funding, and receive lower interest rates and better loan grades. We show that the results are more likely to be driven by the information dissemination channel since social connections are found to enhance the information content of credit grades based on FICO scores, leading to more approvals for higher credit risk borrowers in socially connected areas. Furthermore, all else equal, loans from more socially connected areas are less likely to experience delayed payment or default. The improvement in loan performance is found for all levels of credit grades. Hence, our results support the information dissemination hypothesis that social networks improve information dissemination and reduce information asymmetries.

Then we eliminate endogeneity concerns and show that the positive relationship between social connectivity and loan approval is likely to be causal. Endogeneity may arise from the possibility that social connectivity can be the outcome of some hidden ex ante advantages, qualities, or skills, which may also affect loan approval and loan rates. For example, some unobserved intangible attitudes and beliefs, or similarity with other areas (homophily) can potentially affect both social connectivity and lending outcomes.

To address this endogeneity concern, we employ an exogenous shock that temporarily weakens information diffusion through social connectivity network while not affecting other determinants of lending outcomes: regional flu epidemics. Flu infection causes symptoms such as fever, cough, headache, or fatigue, which directly weaken an investors attention and information processing capabilities. Even if the investors and analysts themselves are not infected, infected family members and relatives can also distract investors attention from work and indirectly slow down information diffusion. The duration of flu symptoms, typically a few days to two weeks, is sufficiently long to impair investors or analysts attention and capability of information processing, which further weaken the capability or efficiency of extracting information from social connections. Although we cannot observe whether each specific person is infected with flu or not, it is safe to assume that investors and their social connections are more likely to be infected if they reside in areas experiencing flu epidemics. We show that regional flu epidemics exhibit significant time-series and cross-sectional variations, which facilitate the implementation of our identification strategies. Our identification strategy allows us to test what will happen to loan origination in an area whose connected areas are experiencing flu activities.

We find that an area's loan approval rate is significantly weakened when its distant but socially connected areas are experiencing flu epidemics. For each one standard deviation increase in the flu measures in friend areas, the effect of social connectivity on loan approval rate is reduced by 40.35% to 67.11% relative to the unconditional mean. The results confirms that information processing and diffusion capacities are weakened when borrowers' friend areas are experiencing flu activities. In sum, our findings suggest that Fintech lending sites allow both borrowers and lenders to make better investment decisions, thereby facilitating more efficient capital allocation.

The rest of the paper is structured as follows. Section 2 introduces the literature and hypothesis development. Section 3 discusses the data and social connectivity measures. Section 4 presents the empirical tests of our hypotheses. Section 5 concludes the paper.

2. Literature and Hypothesis Development

Social networks play an important role in shaping human life and have large effects on social and economic activities. Studies have shown that social interactions are important in information dissemination and preferences formation, thereby influencing the outcomes of trade, labor market, and organizational relations. Using a new measure of social connectedness, Bailey et al. (2017) document that social connectedness is related to a vast array of social and economic characteristics. Although social connectedness is strongly decreasing in geographic distance between counties, the populations of counties with more geographically dispersed online networks are richer, more educated, and have a higher life expectancy. More importantly, Bailey et al. (2017) show that the effects of social connections may spill over to demographic, economic and technological decisions. They find that more socially connected regions have higher trade flows, more cross-region migration and patent citations.

Social interactions can affect economic decisions in a wide range of mechanisms. Hong et al. (2005) find that investors spread information about stocks to one another by word of mouth. Bakshy, et al. (2012) identify the effect of social cues on consumer responses to advertisements, measured in terms of ad clicks and the formation of connections with the advertised entity. Agrawal, Kapur, and McHale (2008) examine how the spatial and social proximity of inventors affects access to knowledge and suggest that although spatial and social proximity both increase the probability of knowledge flows between individuals, the marginal benefit of geographic proximity is greater for inventors who are not socially close. Gee, et al. (2017) show that people find jobs through their social networks using ties of different strengths. Hong et al. (2004) find that stock-market participation is influenced by social interaction.

In addition to promoting the volume of information transfers, social connectivity may enhance the quality of information dissemination, thereby lowering information acquisition costs and reducing information asymmetries that inhibit economic and financial activities across regions. Golub and Jackson (2012) shows that social networks facilitate agreement by lowering the cost of information dissemination and increasing the speed of learning. Han and Yang (2013) analyze a rational expectations equilibrium model to explore the implications of information networks for the financial market, and show that social communication improves market efficiency when information is exogenously endowed.

In addition to the literature showing that closer social connections across regions facilitate communication, reduce cultural differences, cultivate friendships, and enhance mutual understanding, there is a literature that demonstrates how social connections may bias attitudes and preferences (e.g., Rosenblat and Mobius (2004)). People from socially connected regions may build stronger emotional attachments with people from other areas. Occasionally undeserved favorable preferences and opinions may form across socially connected regions as a result of strong mutual recognition. These emotional connections may spill over to financial decisions in that people from more socially connected regions are more recognized by others, and thus are more likely to interact financially and economically (i.e., engage in P2P lending in the context of this paper). For example. Azouzi and Anis (2012) focus on how CEO behavioral biases undermine shareholder value and impact investment decisions by hampering the ability to accurately assess alternatives (optimism and overconfidence) and skewing risk perception (loss aversion).

Cognitive biases engendered by strong social connections may generate persuasion bias that may mislead investment decisions (Rosenblat and Mobius, 2004; De Bondt and Thaler, 1995; Barber and Odean, 2000). DeMarzo et al. (2003) show that persuasion bias, the failure of receivers to account for possible repetition in the messages they hear from others, plays an important role in the process of social opinion formation. Bailey et al. (2017) show that past migration patterns are important determinants of present-day social connectedness. Two regions with the same ancestor are more likely to build strong emotional ties which can affect lending decisions. However, strong attachments may hinder the investor's judgment and interferes with her ability to make rational investment decisions. This may lead to overconfidence and excessive trading (see DeMarzo et al. (2003), Barber and Odean (2000) and Han et al. (2018)).

This paper contributes to the current literature by focusing on the role of social interaction in P2P lending decisions. We examine these decisions from both the demand and supply side. Examining the demand side, both the cognitive bias and information dissemination channels suggest that online social connections encourage loan demand on P2P Fintech sites. That is, we test the following hypothesis against the null hypothesis that online social interactions have no impact on P2P applications:

Hypothesis 1. *Loan applications on online P2P Fintech platforms grow faster in regions whose geographically distant, but socially connected areas experienced larger recent increase in P2P borrowing.*

Further, both cognitive bias and information dissemination channels imply that more extensive social connections enhance loan approval rates either irrationally through behavioral biases or rationally through enhanced screening. This implies the following hypothesis:

Hypothesis 2. *Loan applications from socially connected areas experience increased approval rates.*

The cognitive bias channel implies that socially connected lenders irrationally lend to lower quality borrowers, whereas the information dissemination channel hypothesizes that social connections facilitate the exchange of accurate information used by lenders to screen borrowers. Thus, whether lenders irrationally or rationally favor connected borrowers, we hypothesize that:

Hypothesis 3. *Conditional on being approved, loans from socially connected areas have lower interest rates and receive higher risk grades.*

While both cognitive bias and information dissemination channels have the same implications for the greater likelihood and lower cost of lending across socially connection regions, they have very different implications regarding loan performance. The cognitive bias channel specifies an irrational emotional connection that might favor less creditworthy borrowers from socially connected regions. These loans should experience relatively poor ex post performance (Hypothesis 4a). In contrast, the information dissemination channel specifies that social connections enhance information flows, thereby allowing lenders to base decisions on loan applications from connected regions on accurate information. These loans should experience relatively good ex post performance (Hypothesis 4b). Formally stated:

Hypothesis 4.A. *Ceteris paribus, loans granted to borrowers in socially connected areas experience worse ex post performance than loans originated in areas with less extensive online networks.*

Hypothesis 4.B. *Ceteris paribus, loans granted to borrowers in socially connected areas experience better ex post performance than loans originated in areas with less extensive online networks.*

3. Data and Variables

3.1. Social Connectivity

We adopt the Bailey et al. (2017) measure of social connection index, SCI, which is a county-pair level measure of social connectivity index that uses aggregated and anonymized information from the universe of friendship links between all Facebook users. Facebook, created in 2004 by Mark Zuckerberg, is a social networking site that makes it easy for users to connect and share with their family and friends online. Today, Facebook is the world's largest social network with more than one billion users worldwide. Its success is attributed to its ability to appeal to both people and businesses and its ability to interact with sites around the web. Figure 1 exhibits the findings of a national survey of 1,520 adults conducted in 2016, which shows that Facebook is American's most popular social networking platform by a substantial margin.[4] Nearly eight-in-ten online Americans (79%) used Facebook, more than double the share that used Twitter (24%), Pinterest (31%), Instagram (32%) or LinkedIn

[4] Pew Research Center, November 2016, "Social Media Update 2016."

(29%) in 2016. This implies, on a total population basis (accounting for Americans who do not use the internet at all), that 68% of all U.S. adults were Facebook users in 2016, as compared to 28% on Instagram, 26% on Pinterest, 25% on LinkedIn and 21% on Twitter.

Establishing a friendship link on Facebook requires the consent of both individuals, and there is an upper limit of 5,000 on the number of friends a person can add. Bailey et al. (2017) and Bailey, et al. (2018) find that Facebook's enormous scale, the relative representativeness of its user body, and the fact that individuals primarily use Facebook as a tool to interact with their real-world friends and acquaintances, account for the unique ability of the Facebook social graph to provide a large-scale representation of real-world U.S. friendship networks. In the U.S., Facebook mainly serves as a platform for real-world friends and acquaintances to interact online, with people usually only adding connections on Facebook to individuals whom they know in the real world (Jones, et al. (2013), Gilbert and Karahalios (2009), and Hampton, et al. (2011)).

The Bailey et al. (2017) SCI measure corresponds to the relative frequency of Facebook friendship links between every county-pair. Based on this county-pair-level SCI measure, we calculate an overall connectedness index for each county that measures the total connectedness of a county with all other counties in U.S.:

$$SCI_i = \Sigma_{j=1}^{J} SCI_{i,j}, \tag{1}$$

where $SCI_{i,j}$ is the normalized total number of friendship links between county i and county j constructed by Bailey et al. (2017). The normalized $SCI_{i,j}$ has a maximum value of 1 and minimum value of 0, and relative differences in the SCI correspond to the relative differences in the total number of friendship links. SCI_i is thus the total connectedness of county i with all other counties in the U.S. Let $ShareFrnd_{i,N,j}$ be the share of county i's friends in network N who live in county j:

$$ShareFrnd_{i,N,j} = \frac{SCI_{i,j}^{N}}{SCI_{i}^{N}} \tag{2}$$

Network N can be the broadest network that includes all Facebook connections or some other subnetworks such as out-of-commuting zone connections or out-of-state connections. Higher $ShareFrnd_{i,N,j}$ means that more of county i's friends live in county j.

Note that $SCI_{i,j}$ is not adjusted for local population, so it is biased up in counties with higher population. Bailey et al. (2017) thus construct an alternative connectedness measure which is scaled by local population:

$$RSCI_i = \frac{\Sigma_{j=1}^{J} SCI_{i,j}}{Pop_i} = \frac{SCI_i}{Pop_i} \tag{3}$$

where Pop_i is the population in county i. Pop_i can be replaced with the number of Facebook users in county i, and Bailey et al. (2017) show that the two choices lead to very similar effects. In the rest of this paper, we will use $RSCI_i$ as our main measure of social interaction between region i and the rest of the country.

3.2. Peer-to-Peer (P2P) Lending

We obtain personal loan data from the largest P2P lending platform, LendingClub, for the period of 2010 to 2017[5]. LendingClub provides detailed information regarding borrower creditworthiness, including FICO scores, debt-to-income ratio, home ownership, employment length, etc. P2P lending platforms offer lending-based crowd-funding for consumers and small businesses where multiple lenders lend small sums of money online to consumers or small businesses with the expectation of periodic repayment. LendingClub launched the first online P2P lending platforms in the US in 2007, followed by other companies such as Upstart, Funding Circle, CircleBack Lending or Peerform. In December 2014, LendingClub became the first publicly traded online P2P lending platform in the U.S. after its Initial Public Offering on the New York Stock Exchange. As of September 2017, LendingClub has

[5] Our sample starts from 2010 because loans originated before 2010 only account for less than 0.5% of total number of loans originated between 2007, the year of its establishment, and 2017.

intermediated $28 billion of loans, followed by Prosper, which issued $10 billion of loans.

Consumer loan amounts vary between a minimum loan of $500 and a maximum loan of $35,000 ($300,000 for businesses). The platform funds various types of projects ranging from credit card debt consolidation to home improvement, short-term and bridge loans, vehicle loans or engagement loans. As in many other two-sided markets (see Rysman (2009)), online lending marketplaces try to attract two different groups of users, namely borrowers and investors, by choosing an appropriate structure of fees that depends on the magnitude of cross-network externalities. On the borrower side of the market, online lending marketplaces compete with banking institutions, credit unions, credit card issuers and other consumer finance companies. Platforms claim that their prices are lower on average than the interest rates and fees that consumers pay on outstanding credit card balances or unsecured installment loans funded by traditional banks. Online lending marketplaces perform some of the traditional screening function of banks by defining various broad criteria that must be met by borrowers. However, online lenders screen the individual loan applications individually using their own information sources.

During the sample period, there are 19,769,249 loan applications, of which 1,705,423 applications were approved, equivalent to an approval rate is 8.63%. Of the 19,769,249 loan applications, we drop 9,639,920 applications whose FICO scores are missing. LendingClub requires a minimum FICO score of 660 for an application to be considered, and indeed we observe that during our sample period, the lowest FICO score for an approved loan is 660. Since FICO scores perfectly predict rejections for applications with FICO scores below 660, we only keep those ones with FICO equal to or higher than 660 to examine the effect of social interaction on loan approval. This leads to a sample of 4,607,275 applications. We then merge LendingClub data with social connectivity data, census data, bank presence and competition data, and employment data, and obtain a sample of 4,097,363 applications, of which 1,489,527 were approved.

Social connectivity can be relevant for P2P lending activities for a few reasons. First, LendingClub was initially launched on Facebook in 2006 as one of Facebooks first applications. The social networking advantage of Facebook allows LendingClub to leverage trust of enabling lenders to find borrowers within shared network. LendingClub developed an algorithm called "LendingMatch" which finds relationships between borrowers and lenders based on geography, education, profession, or connectedness within a given social network and then presents lenders with diversified loan portfolios reflecting these relationships, while not giving lenders direct access to the borrowers' Facebook profile. This matching mechanism underscores a founding principle of Lendingclub - they believe that bringing lenders and borrowers together in a marketplace that leverages existing communities and relationships will make for a better experience for all participants. Matching lenders with compatible borrowers helps loans get funded, while letting borrowers know that members of their community supported their financial needs increases the overall performance of the loan, as borrowers would be less likely to default. Operating entirely via Facebook, LendingClub grew to one of the top 10 Facebook apps within weeks following its launch. LendingClubs growing success demonstrates the potential of Facebook as a sales and finance platform.

Second, although LendingClub has detached from Facebook and developed into a full-scale peer-to-peer lending company, the impress of social interactions remain on its lending activities. LendingClub and Facebook, as well as other social media, share a similar customer base. The main group of Facebook users are young people who are also the largest group of borrowers on LendingClub. For example, Figure 2 shows the distribution of employment length for LendingClub loan applicants between 2016 and 2017, as compared to the age distribution of Facebook users according to a survey conducted in January 2018. Figure 2 demonstrates that most of the LendingClub borrowers have employment length shorter than 2 years, indicating that they are mainly young people. Similarly, Facebook users are dominated by those in the 18-29 age bracket. Thus, even though we are not sure about whether a specific Facebook user is also a LendingClub borrower, we can generally infer that the social connectivity index by Facebook is relatively representative for the population of LendingClub borrowers. Further, both LendingClub and Facebook target individuals (either as consumers or small business owners), thereby eliminating any distortions from institutions and their online networks. Thus, the social connectivity data from Facebook provide valuable insights into the P2P online loan market.

The social interaction measure of Bailey et al. (2017) is at the county-pair level, whereas the geographic locations of loan borrowers are identified by state and 3-digit zip code on LendingClub. In U.S., each county can contain multiple different 3-digit zip codes, and a 3-digit zip code can reside different counties. To match counties with zip codes in each state, we combine all counties containing a specific 3-digit zip code and treat them as a single area. For example, six counties (Fayette, Lee, Colorado, Austin, Bastrop, Gonzales) in Texas contain the 3-digit zip code 789. In this case, we combine these six counties and treat them as the area corresponding to the 3-digit zip code 789. The connectedness between this area and other counties are then calculated. Our sample contains 911 unique 3-digit zip codes and 3,136 unique counties and county equivalent entities, covering the vast majority of the U.S. Our sample covers the periods between 2010 to 2017. Table 1 provides a list of variables used with definitions of each variable.

LendingClub assigns each approved loan a loan grade (A, B, C, D, E, F, or G) based on its internal credit evaluation methodologies. Grade A loans are the least risky and Grade G loans are most risky. Each grade is further categorized into five subgrades (e.g., Grade A has subgrades of A1, A2, A3, A4, A5) based on loans credibility. Thus, there are in total 35 subgrades of loans. We create a *LoanGrade* variable with value of 35 for A1 and decrease its value by 1 for each subgrade below A1.

3.3. Control Variables

We control for a wide range of control variables that are potentially related to social connectivity and also affect borrowing choices and lending outcomes. A large literature has documented that many individual characteristics, including racial identity, gender, education are associated with social connectivity (McPherson, Smith-Lovin, and Cook, 2001). We obtain local social, economic, and demographic characteristics data from the 5-year estimates of the American Community Survey. Demographic characteristics include population, share of senior population (65+ years old), share of white population, share of female population. Economic characteristics include income per capita, unemployment rate, share of labor force in manufacturing industry, share of labor force working in information industry, number of bank branches per capita, amount of bank deposit per capita and local bank deposit concentration as measured by the Herfindahl-Hirschman Index (HHI). Social characteristics include the share of people with higher than a high school education in the population aged 25 years or older.

3.4. Descriptive Statistics

Table 2 presents the LendingClub summary statistics for loan applications and originations during the sample period of January 2010 through December 2017. Panel A presents the summary statistics for the loan application sample, including both approved and rejected loans, and Panel B presents summary statistics for approved loans only. The median loan amount of the application sample is $10,000, whereas the median amount of approved loans is $12,000. The median FICO score in the application sample is 652, whereas the median FICO in the approved loan sample is 692. The median debt-to-income (DTI) ratio in the application sample is 20%, and the median DTI ratio in the approved loan sample is 17.96%. The median employment length is 1 year in the application sample, and 6 years in the approved loan sample. The annual income of borrowers in the approved loan sample is $65,000.

The average loan rate is 13.2% in the origination sample. In sum, approved loans have larger amounts, higher FICO scores, lower debt-to-income ratios, and longer employment lengths.

The summary statistics shows that the average $RSCI$ is slightly higher in the origination sample than in the application sample, implying a positive relationship between loan approval and online social connectivity. In the application (origination) sample, the $RSCI$ measure has a mean value of 0.132 (0.133) and standard deviation of 0.040 (0.039). A t test significantly rejects the null hypothesis that the mean values of $RSCI$ in the application sample and approved sample are the same. Table 3 presents an area-level correlation matrix for variables included in the origination sample. It shows that loan rates are positively correlated with funded amount, loan term, and debt-to-income ratio, and are negatively correlated with loan grade, employment length, FICO scores, and annual income. Interest rates are positively correlated with $RSCI$ measures, but statistically insignificant. Borrower annual income is also positively related to larger funded amount, longer loan term, longer employment length, and higher FICO score.

To assess the performance characteristics of the loans, we take a snapshot of a recent sub-sample of loans originated in 2015 and report the loan status within the first 24 months after origination[6]. Panels C of Table 2 present the summary statistics of monthly loan status and Panel D shows the number of normal loans (i.e., loans that never experienced any late payment, default, or charge-off within 24 months after origination.) and abnormal (i.e., loans that experienced default, charge-off, or at least one late payment within 24 months after origination.) loans. It shows that 66,292 out of 421,095 loans originated in 2015 experience at least one form of bad performance, with the most frequent form of bad performance being late payment by 31-120 days.

[6]LendingClub loans have either 3-year or 5-year terms that are generally amortized over 36 or 60 monthly payments, respectively. For loans originated between 2010 and 2017, the median loan term is 36 months.

4. Empirical Strategies and Results

4.1. The Effects of Social Interactions on Borrowing Decisions

To examine the demand side and test Hypothesis 1, we examine the relationship between loan applications and social connectivity. We adopt the methodology of Bailey et al. (2018) and test whether there is a causal relationship between P2P borrowing decisions and past P2P borrowing experiences of connected friends in geographically distant regions. We define region i's friends' P2P borrowing experience as:

$$FrndExp^{N}_{i,t_1,t_2} = \sum_j ShareFrnd_{i,N,j} \times Exp_{j,t_1,t_2}. \tag{4}$$

where Exp_{j,t_1,t_2} is region j's P2P borrowing history during the period between time t_1 and t_2. Borrowing history is measured using four different metrics: loan application growth between t_1 and t_2 ($\Delta FrndApp_{i,t_1,t_2}$), loan origination growth between t_1 and t_2 ($\Delta FrndOrg_{i,t_1,t_2}$), changes in approval rate ($\Delta FrndRate_{i,t_1,t_2}$), and the average approval rate across t_1 and t_2 ($FrndRate_{i,0}$). Both loan application and origination measures are scaled by local population.

We then analyze how socially connected friends' P2P experience influence ones borrowing decisions by regress an area's P2P borrowing activities on the past borrowing activities of its connected regions. Specifically, we estimate the following regression:

$$\Delta Application_{i,t_2,t_3} = \beta FrndExp^{N}_{i,t_1,t_2} + \gamma \mathbf{X}_{i,t_2} + \epsilon_{i,t_2,t_3} \tag{5}$$

where $\Delta Application_{i,t_2,t_3}$ is changes in loan applications for area i between time t_2 and t_3, X is a wide range of demographic, economic, and social characteristics that may also affect the P2P borrowing decisions and outcome, and t_1, t_2, and t_3 represent year 2010, 2013, and 2016, respectively. We focus on area i's broadest network so N represents all of area i's

Facebook friends. Essentially, we predict a region's borrowing activities during the period of 2013-2016 (response period) using its connected friends' borrowing history during the years 2010 to 2013. This specification of time periods allows us to alleviate the concern that the relationship of P2P activities between a county with its friend counties can be driven by common macroeconomic variations across these counties.

Equation (5) may still involve endogeneity issues if the past borrowing experience of connected friends is driven by a region's own borrowing experiences. For example, if most of the the connected friends are located in geographically approximate areas, then common economic variations may drive a spurious relation between a county's borrowing activities and its friends borrowing experiences. To address this concern, we follow Bailey et al. (2018) and employ an instrumental variables (IV) approach, where we instrument for an area's friend experience with the past borrowing history of its geographically distant friends, such as out-of-state connections. This way, we are able to remove, form the friends' borrowing experience variable, the underlying economic conditions that may impact nearby regions jointly. Therefore, we replace the friends' borrowing experience variable in equation (5) with an instrumented variable and estimate the following two-stage IV regression:

$$FrndExp^{All}_{i,t_1,t_2} = \beta^{FS} FrndExp^{OutState}_{i,t_1,t_2} + \delta \mathbf{X}_{i,t_1} + s_k + \epsilon_{i,t_1,t_2} \tag{6}$$

and

$$\Delta Application_{i,t_2,t_3} = \beta^{IV} \widehat{FrndExp^{All}_{i,t_1,t_2}} + \gamma \mathbf{X}_{i,t_2} + s_k + \epsilon_{i,t_2,t_3}, \tag{7}$$

where $FrndExp^{All}_{i,t_1,t_2}$ is the average borrowing history of an area's all connections, $FrndExp^{OutState}_{i,t_1,t_2}$ is the average borrowing history of its out-of-state connections, and $\mathbf{X}_{i,t_1}$ is a vector containing local demographic, economic, and social characteristics that may also affect local P2P borrowing decisions. It includes local bank branches per one thousand population ($Branch$), deposit per capita ($Deposit$), local banking competition ($HHI_{i,t}$), local population ($Population$, in millions), share of white people ($White$) in local population, share of female ($Female$) in local population, share of people with higher than high school education ($Education$)) in the population aged 25 years or older, share of people older than 65 ($Senior_{i,t}$), income per capita ($Income$), unemployment rate ($Unemployment$), share of labor force working in the manufacturing industry ($Manufacuture$), and share of labor force working in the information industry ($Information$). We include state fixed effects (s_k) to control for with-in state patterns of P2P demand, and cluster standard errors at the state level.

Table 4 presents the results for Equation (6), the first stage regression of the IV approach for 2010 to 2013, with Column (1) to (4) reporting the results on loan application, origination, growth in approval rates, and average approval rates. β^{FS} is positive and significant across different measures of borrowing history in connected regions, indicating that friends' P2P activity across all connected regions is positively correlated with friends' P2P history in out-of-state connected regions.

Table 5 shows the results for Equation (7), the second stage regression of the IV approach, where the fitted value of instrumented variable from the first stage is used as the main independent variable. Table 5 shows that region i's loan applications experience larger increases if connected regions had larger past increases in loan applications. Economically, on average, if the growth of applications per one thousand people increases by 10 applications between 2010 to 2013 in friend areas, then the growth of applications per one thousand people increases by 2.8 applications between 2013 to 2016. The results are robust if we replace friend areas past loan application experiences with loan origination, average approval rates, or increases in approval rates. The results imply that borrowers are influenced by their social network friends' borrowing activities.

4.2. Loan Approval and SCI

To test Hypothesis 2, we utilize a logit regression using loan application outcome as the dependent variable and the borrower county's social interaction index as the key explanatory variable. Specifically, we estimate the following regression at the loan level:

$$Approval_{i,k} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k} \tag{8}$$

where i indexes region and k indexes loan, $Approval_{i,k}$ is 1 for approved loans and 0 for rejected loans, and $RSCI_i$ is the relative social connectedness index of region i.[7] $\boldsymbol{X}_k$ is a vector of loan, borrower, demographic, economic, and social characteristics. s_i indexes state fixed effects and q_k indexes application quarter fixed effects. LendingClub's rejected loans database contain fewer borrower characteristics variables than the approved loan database, which also provide borrower income and loan maturity, so the variables in the regressions are limited to variables included in both approved and rejected databases. Based on Hypothesis 2, we expect β to be significantly positive.

Table 6 presents the results for Equation (8). The coefficient on $RSCI$ is positive and statistically significant at the 1% level, consistent with the prediction of Hypothesis 2. Taking Column (3) of Table 6 as an example, starting from the mean approval rate of 36.34% (standard deviation of 48.22%), a one standard deviation increase in $RSCI$ is associated with increase in the likelihood of loan approval by 0.549%. The signs of coefficients on control variables are as expected. Borrowers with higher FICO Score, lower DTI, and longer employment length have greater likelihood of receiving loan approvals. Loans with larger amounts are less likely to be approved.

Next, we examine whether the relationships between social connectivity and loan approval are different for borrowers of different credit worthiness. If social interactions help promote

[7] Approved loans that cannot be sufficiently funded are also categorized as rejected in the LendingClub database.

the information dissemination between connected areas, then both favorable and unfavorable information about the borrowers can be disseminated to potential lenders, and it is possible that credit worthy borrowers benefit more from broader social interactions than less credit worthy ones. Thus, we divide the sample into three sub-samples of equal size based on FICO scores and re-estimate Equation (8).

Table 7 reports the results. Social connectivity is positively correlated with loan approval in all three sub-samples at either 1% or 5% significance levels, while the positive relationships are stronger for high FICO score groups than for low FICO groups. Starting from the mean, each one standard deviation increase in $RSCI$ is associated with 0.45%, 0.56%, and 0.64% increases in likelihood of approval for the low, medium, and high FICO groups, respectively. This implies that while social connectivity improve funding outcomes for borrowers of all levels of credibility, it benefits creditworthy borrowers more than it does for less creditworthy ones.

In addition to analyzing the loan-level approval probability, we examine the approval rate by conducting an area level analysis. We aggregate loan applications to the region-quarter level and calculate the quarterly approval rate of each region. The quarterly approval rate is defined as the total number (or dollar amount) of originated loans divided by the total number (or dollar amount) of all loan applications. Note that the approval rate is an approximate estimate of the true lender approval rate, since some applications were withdrawn by the borrower or the loans were not funded to the required amount. We then estimate the Tobit regressions of the approval rate on social interaction and a range of control variables:

$$ApprovalRate_{i,t} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,t} + s_k + q_t + \epsilon_{i,k} \tag{9}$$

Table 8 presents the results of Equation (9). The coefficient on $RSCI$ is positive and significant, indicating that loans borrowed from more socially connected areas are more likely to get approved, which confirms the results in Table 6.

Some areas experienced very low volume of applications during some periods and experienced either 0% or 100% approval rates, in which case it is hard to draw economic and statistical inference from the approval rates. To further confirm the robustness of the relationship between social connectivity and loan approval, we drop those area-month observations with either 0% or 100% approval rates, which constitute about 5% of all observations. Dropping observations at lower and upper bounds allows us to estimate 9 using linear estimation method. 9 shows the results of linear regressions, which exhibit similar results as non-linear regressions.

4.3. Flu Epidemics as Exogenous Shocks to Social Connectivity Intensity

Endogeneity may arise from the possibility that social connectivity can be the outcome of some hidden ex ante advantages, qualities, or skills, which may also affect loan approval rates, interest rates, and likelihood of default. Although we have controlled for a range of observable local social, economic, and demographic characteristics that can explain both social connectivity and lending outcome, there still could be some other unobserved characteristics, such as intangible attitudes and beliefs, or similarity with other areas (homophily), that can potentially affect both social connectivity and lending outcomes.

To address this endogeneity concern and isolate the causal effect of social connectivity on lending outcome, we employ an exogenous shock that temporarily weakens information diffusion through social connectivity networks while not affecting other determinants of lending outcomes: regional flu epidemics. Flu infection causes symptoms such as fever, cough, headache, or fatigue, and investors' attention and information processing capabilities are directly weakened when they are infected with the flu. Even if the investors and analysts themselves are not infected, infected family members and relatives can also distract investors' attention from work and indirectly slow down information diffusion. Infected colleagues can also reduce capability of information processing due to weaker collaboration and support. The duration of flu symptoms, typically a few days to two weeks, is sufficiently long to impair investors' or analysts' attention and capability of information processing, which further weaken the capability or efficiency of extracting information from social connections. Although we cannot observe whether each specific person is infected with flu or not, it is safe to assume that investors and their social connections are more likely to be infected if they reside in areas experiencing flu epidemics.

We obtain flu epidemic data spanning 2010 to 2017 from the Outpatient Illness and Viral Surveillance database provided by the Center for Disease Control and Prevention (CDC) and WHO/NREVSS. We employ two measures of flu epidemics: one is 'the percentage of patient visits to healthcare providers for influenza-like illness' (ILI) and the other is 'the percentage of flu tests with positive results' (POS), both of which are observed at the state level on weekly basis. We aggregate the measures into monthly values to keep consistency with our measurement of loan origination.

Figures 3 and 4 present the statistical properties of the two measures of flu epidemics. Figure 3 shows the monthly time-series of the flu measures. From the figure we observe that flu epidemics usually occur in winter and spring seasons, highlighting the importance of controlling for seasonality in empirical tests. The figure also shows that the duration and severity of flu epidemics vary a lot across years, which provides enough time-series variations for our identification strategies. Figure 4 shows the relationship between the monthly national-level value of flu measures and monthly cross-sectional variations in flu measures across all U.S. states (each dot represents a month). We can observe the severity of flu epidemics is positively associated with variations in flu activities across U.S. states. This ensures enough cross-sectional variations for our identification strategies.

Our identification strategy allows us to test what will happen to loan originations in an area whose connected areas are experiencing flu activities. For example, assume Area A and

Area B are a pair of well-connected areas. When Area B is affected by a serious flu epidemic that lasts for one month, the social, economic, demographic fundamentals of Area A are not affected, but social connection of Areas A with Area B is temporarily weakened. We then test what will happen to the loan origination in Area A during this flu period of Area B, relative to other areas with friend areas that are not affected.

To measure the aggregate severity of flu activities in all area i's connected areas, we first construct the following flu severity measure for area i that is customized for each area's geographic distribution of online social connections:

$$FluImpact_{i,t} = \Sigma_{j=1}^{J} RSCI_{i,j}^{OS} \times FluMeasure_{j,t}, state(j) \neq state(i) \tag{10}$$

where i, j, t respectively denote own area (Zip Code), friend state, and time (month), where j excludes the state where area i is located in. $FluMeasure$ is one of two state-level measures of flu epidemics, ILI or POS. $RSCI_{i,j}^{OS}$ represents the relative social connectivity index between area i and state j, where i and j have to be in different states. Higher $FluImpact$ indicates that area i's friend areas are more severely affected by the flu during month t.

Based on the premise that investors, analysts, and their family and friends are more likely to be infected with flu in flu-affected areas, we identify regional peaks of flu pandemics that are severe enough to impact investors. We define regional peaks of flu pandemic as those months when the flu measures (ILI, POS, or both) are higher than both the time-series average for the state and the concurrent average flu measures across all U.S. states. Thus, the $FluImpact$ variable can be revised to:

$$FluImpact_{i,t}^{P} = \Sigma_{j=1}^{J} RSCI_{i,j}^{OS} \times FluPeak_{j,t}, state(j) \neq state(i) \tag{11}$$

where i, j, t denote own area (Zip Code), friend state, and time (month), where j excludes the state where area i is located in. $RSCI_{i,j}$ represents the relative social connectivity index between area i and state j. $FluPeak$ is a dummy variable with the value one if the flu measure (ILI, POS, or both) is simultaneously higher than the time-series average for the state and the concurrent average across all U.S. states, and zero otherwise. Thus, $FluImpact^{P}_{i,t}$ measures the extent to which an area's friends are affected by serious flu pandemics.

To analyze how loan approval rates change when flu epidemics shock an area's online social connections, we estimate the following fixed effects regression:

$$ApprovalRate_{i,t} = \beta_1 RSCI^{OS}_i + \beta_2 FluImpact_{i,t} + \gamma X_{i,t} + s_i + \theta_t + \epsilon_{i,t} \tag{12}$$

where i and t denote area and time (month). $ApprovalRate_{i,t}$ is the loan approval rate (total origination count or amount divided by total application count or amount). $X_{i,t}$ is a vector of loan, borrower, demographic, economic, and social characteristics. s_k indexes state fixed effects and $_t$ indexes application time fixed effects. $FluImpact_{i,t}$ is the flu impact measure we constructed in Equation (10) and can be either $FluImpact^{ILI}_{i,t}$ or $FluImpact^{POS}_{i,t}$ depending on the flu measure used. Flu Impact can further be replaced with $FluImpact^{P}_{i,t}$ calculated from Equation (11).

Although we are testing how a weakened social connectivity caused by flu epidemics affects loan origination, note that in Equation () there is no interaction term between $RSCI$ and $FluImpact$. This is because Flu Impact, by construction, already includes an interaction between social connection and flu measures, and the coefficient on it measures the marginal impact of social connectivity on loan origination that is driven by flu epidemics. In $X_{i,t}$, we also control for the flu measures for all area i's connected areas to control for their direct effect .

Table 10 presents the estimation results for Equation (4.3). Panel A uses the $FluImpact$ measure constructed in Equation (10) and Panel B uses the $FluImpact^{P}_{i,t}$ constructed in Equation (11). We observe two results in Panels A and B. First, the positive relationship between $RSCI$ and loan approval rate persists in both sets of tests. Second, the coefficients on $FluImpact$ measures are negative and significant at the 1% level across all specifications, indicating that the positive relationship between $RSCI$ and loan approval rate is weakened if the friend areas are experiencing flu epidemics. For example, the coefficient estimate on $FluImpact$ in Column (1) of Panel A is -6.679, which is statistically significant at the 1% level. This result suggests that, when an area's friend areas are experiencing flu activities (i.e., the ILI measure increases by one standard deviation in all the friend areas), the effect of social connectivity on loan approval rate is reduced by 67.11%, relative to the unconditional mean. Similarly, when using the measure POS as reported in Column (3), the effect of social connectivity on loan approval rate is reduced by 40.35% for each one standard deviation increase in flu measure. Similar results are shown in Panel B.

One concern about Equation (4.3) is that the values of the dependent variable, the approval rate, are economically restricted to lie in a certain interval between 0 and 1. However, OLS regressions like Equation (4.3) do not account for this fractionality and assume that the dependent variable can take on every negative or positive real number, which leads to biased estimation (Heckman, 1979; Elsas and Florysiak, 2015). To resolve this concern, we re-estimate Equation (4.3) using a double boundaries Tobit estimation (with boundaries 0 and 100). Table 11 reports the results for the Tobit models. The results do not qualitatively change in the Tobit model estimations.

4.4. Loan Pricing and SCI

To investigate the relationship between loan pricing and social interaction (Hypothesis 3), we focus on approved loans and estimate fixed effects regressions using either the loan interest rate or LendingClub assigned loan grades as the dependent variable. We start with the following fixed effects regressions:

$$InterestRate_{i,k} = \alpha + \beta RSCI_i + \gamma \boldsymbol{X}_{i,t} + s_k + q_t + \epsilon_{i,k} \tag{13}$$

where i indexes county and k indexes loan. $InterestRate_{i,k}$ is the interest rate for loan k granted to a borrower in county i. $\boldsymbol{X}_k$ is a vector of loan, borrower, demographic, economic, and social characteristics. Based on Hypothesis 3, we expect β to be negative and significant.

If social connectivity helps borrowers obtain better loan terms, then loans from higher social connectivity areas should be assigned better loan grades, thus we replace the $InterestRate_{i,k}$ with $LoanGrade_{i,k}$ and run the following regression:

$$LoanGrade_{i,k} = \alpha + \beta RSCI_i + \gamma \boldsymbol{X}_{i,t} + s_k + q_t + \epsilon_{i,k} \tag{14}$$

Table 12 presents the results for Equation (13). Consistent with Hypothesis 3 that predicts a negative β, the coefficient on $RSCI$ is negative and significant at the 1% level. Economically, using the full regression results in Column (3), for each one standard deviation increase in $RSCI$, interest rate decreases by 7 basis points. With regard to the control variables, borrowers with higher FICO score, higher annual income, lower DTI, and longer employment length obtain better interest rates from lenders. Table 12 also shows that LendingClub charges higher interest rates on larger funded amounts and longer term loans. Loan rates are also lower in areas with higher populations, more bank branches per one thousand population and higher deposits per capita. Areas with higher shares of white people, better education, and lower unemployment rates also get lower interest rates.

Columns (3) and (4) of Table 12 present the results for Equation (14). Consistent with Hypothesis 3, the coefficient on $RSCI$ is positive and significant at the 1% level. Economically, using the full regression results from Column (3), for each one standard deviation increase in $RSCI$, loan grade increases by 0.1 sub-grade out of a total of 35 grades. Both the borrower financial profile and loan characteristics are important elements in evaluating Loan Grades. Loans initiated by good borrowers with higher FICO score, higher annual income, lower DTI, and longer employment length are assigned better grades. Larger loans with higher funded amounts tend to get worse grades. Short term loans get better grades compared to long term loans. Loans to borrowers in areas with more bank branches and deposits per capita, a higher share of white population, better education, higher income, lower unemployment rate all are assigned better loan grades. Untabulated results also show that the results are robust across low, medium and high FICO score sub-samples.

In sum, Table 12 provides evidence on the pricing implication of social interactions. That is, more socially connected areas receive lower loan rates and higher loan grades.

4.5. Loan Performance and SCI

In this section, we differentiate between the cognitive bias and information dissemination hypotheses using ex post loan performance. That is, if the positive relationship between social connections and lending activity is driven by the emotional and cognitive bias channel, then we should observe more defaults and late payments for loans with the same ex ante loan grade and interest rate designations (Hypothesis 4.A). Alternatively, if social connectedness reduces information asymmetries between borrowers and lenders, then loans originated in more connected areas should experience lower default rates or fewer late payments (Hypothesis 4.B). We utilize all loans originated during our sample period and investigate their 24-month performance. LendingClub provides the performance data for all loans originated. For each loan in each month, the performance database provides loan status, beginning balance, principal and interest paid, amount due, amount paid.

The loan status at the end of each month after loan origination is one of the followings: *Issued*, *Current*, *In Grace Period*, *Late (16-30 days)*, *Late (31-120 days)*, *Default*, *Charged Off*, and *Fully Paid*. A loan is at the status of *Issued* in the month of origination. The status is *Current* if a monthly payment is made before due date. When a borrower misses a loan payment, the loan status changes from Current to Late.

LendingClub gives all borrowers a 15-day grace period to make payments with no penalty. Thus if a borrower misses a monthly payment and the end of that month is within 1-15 days after the due date, then the status becomes *In Grace Period* for that month. If borrowers made a payment within the first 15 days after their due date, they would not accrue a late fee. Before Feb 24, 2017, if a borrower pays between 1-15 days past their due date, then any interest accrued during grace period would be waived. This grace period interest waiver policy was eliminated after Feb 24, 2017. If LendingClub did not receive payment during the grace period, it may charge a late fee. If a payment was not received within 16-30 days after the due date, the status becomes *Late (16-30 days)*. If a payment was not received within 31-120 days after the due date, the status becomes *Late (31-120 days)*. When a borrower misses several payments, the loan will enter *Default* status. When there is no longer a reasonable expectation of further borrower payments, the loan will be *Charged Off* by the platform. The status becomes *Fully Paid* if the loan is fully paid off. A loan is defined as normal loan if its status stays as *Issued*, *Current*, *Fully Paid*, or *In Grace Period* during the 24-month period after loan origination. A loan is defined as an abnormal loan if its status falls into *Late (16-30 days)*, *Late (31-120 days)*, *Default*, or *Charged Off* at least once during the 24-month after loan origination.

To test the performance of loans originated in different social connectedness areas. We run a set of logit regressions:

$$BadPerformance_{i,k} = \alpha + \beta RSCI_i + \gamma \boldsymbol{X}_{i,t} + s_k + q_t + \epsilon_{i,k} \tag{15}$$

where i indexes area and k indexes loan. $BadPerformance_{i,k}$ equals 1 if at least one of the following statuses occurs on loan k within 24 months after loan origination: *Late (16-30 days)*, *Late (31-120 days)*, *Default*, *Charged Off*, and 0 otherwise.

Hypothesis 4.A predicts that the coefficient on β is expected to be positive. Alternatively, Hypothesis 4.B predicts that the coefficient on $RSCI$ is expected to be negative. Panel A of Table 13 presents the results for Equation (15). Consistent with Hypothesis 4.B, β is negative at the 1% level. Economically, using the full regression results in Column (3), starting from median level of 3.42%, for each one standard deviation increase in $RSCI$, the likelihood of bad performance decreases by 0.15%. Moreover, Panel B of Table (13) shows an improvement in performance across all credit qualities from marginal increases in social connectivity, such that the improvements are higher for high FICO borrowers. Unreported tests show that the coefficients on $RSCI$ across FICO scores are significantly different from each other at the 1% significance level. Hence, our results support the information dissemination channel, hypothesizing that online social networks improve information dissemination and reduce information asymmetry.

5. Conclusion

In this paper, we use a novel measure, the social connectivity index (SCI), to examine the effects of social connections on the demand for and supply of consumer and small business loans on peer-to-peer (P2P) FinTech sites such as LendingClub. We find that P2P loan demand increases when geographically distant, but socially connected areas have large amounts of past P2P borrowing activity. This suggests that social networks provide a mechanism for propagation across markets of awareness about and recognition of alternative sources of borrowing. Loans originated in high SCI areas are more likely to be approved and funded. Both approval rates and quality (as measured by assigned loan grades and interest rates) are higher the greater an area's aggregate social connectivity. We then examine whether the results are driven by an irrational emotional connection channel or by a rational information asymmetry channel. We find that loans originated in high SCI areas are less likely to experience late payment or default, which suggests that social connectivity facilitates P2P lending through a channel of reducing information asymmetries and uncertainty, thereby assisting rational decision making. In sum, our findings suggest that social connectivity plays an important role in affecting individual economic decisions in the context of FinTech P2P on-line lending.

References

Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? evidence from patent data. Journal of Urban Economics 64, 258–269.

Azouzi, M., Anis, J., 2012. Ceo emotional bias and investment decision, bayesian network method. Management Science Letters 2, 1259–1278.

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

Bailey, M., Cao, R. R., Kuchler, T., Stroebel, J., Wong, A., 2017. Measuring social connectedness. Tech. rep., National Bureau of Economic Research.

Bakshy, E., Eckles, D., Yan, R., Rosenn, I., 2012. Social influence in social advertising: Evidence from field experiments. In: *Proceedings of the 13th ACM Conference on Electronic Commerce*, ACM, pp. 146–161.

Barber, B. M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. The Journal of Finance 55, 773–806.

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

Cohen, L., Frazzini, A., Malloy, C., 2010. Sell-side school ties. The Journal of Finance 65, 1409–1437.

Coval, J. D., Moskowitz, T. J., 1999. Home bias at home: Local equity preference in domestic portfolios. The Journal of Finance 54, 2045–2073.

De Bondt, W. F., Thaler, R. H., 1995. Financial decision-making in markets and firms: A behavioral perspective. Handbooks in Operations Research and Management Science 9, 385–410.

DeMarzo, P. M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. The Quarterly Journal of Fconomics 118, 909–968.

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

Elsas, R., Florysiak, D., 2015. Dynamic capital structure adjustment and the impact of fractional dependent variables. Journal of Financial and Quantitative Analysis 50, 1105–1133.

Ge, R., Gu, B., Feng, J., 2017. Borrowers self-disclosure of social media information in p2p lending. In: *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., Fowler, J. H., 2017. The paradox of weak ties in 55 countries. Journal of Economic Behavior & Organization 133, 362–372.

Ghosh, P., Ray, D., 2016. Information and enforcement in informal credit markets. Economica 83, 59–90.

Gilbert, E., Karahalios, K., 2009. Predicting tie strength with social media. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, pp. 211–220.

Golub, B., Jackson, M. O., 2012. How homophily affects the speed of learning and best-response dynamics. The Quarterly Journal of Economics 127, 1287–1338.

Hampton, K., Goulet, L. S., Rainie, L., Purcell, K., 2011. Social networking sites and our lives .

Han, B., Hirshleifer, D., Walden, J., 2018. Social transmission bias and investor behavior. Tech. rep., National Bureau of Economic Research.

Han, B., Yang, L., 2013. Social networks, information acquisition, and asset prices. Management Science 59, 1444–1457.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Hong, H., Kubik, J. D., Stein, J. C., 2004. Social interaction and stock-market participation. The Journal of finance 59, 137–163.

Hong, H., Kubik, J. D., Stein, J. C., 2005. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. The Journal of Finance 60, 2801–2824.

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

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., Fowler, J. H., 2013. Inferring tie strength from online directed behavior. PloS one 8, e52168.

Kelly, M., O'Grada, C., 2000. Market contagion: Evidence from the panics of 1854 and 1857. American Economic Review 90, 1110–1124.

Massa, M., Simonov, A., 2005. Behavioral biases and investment. Review of Finance 9, 483–507.

McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. Annual Review of Sociology 27, 415–444.

Rosenblat, T. S., Mobius, M. M., 2004. Getting closer or drifting apart? The Quarterly Journal of Economics 119, 971–1009.

Rysman, M., 2009. The economics of two-sided markets. Journal of Economic Perspectives 23, 125–43.

Shiller, R. J., Pound, J., 1989. Survey evidence on diffusion of interest and information among investors. Journal of Economic Behavior & Organization 12, 47–66.

Shive, S., 2010. An epidemic model of investor behavior. Journal of Financial and Quantitative Analysis 45, 169–198.

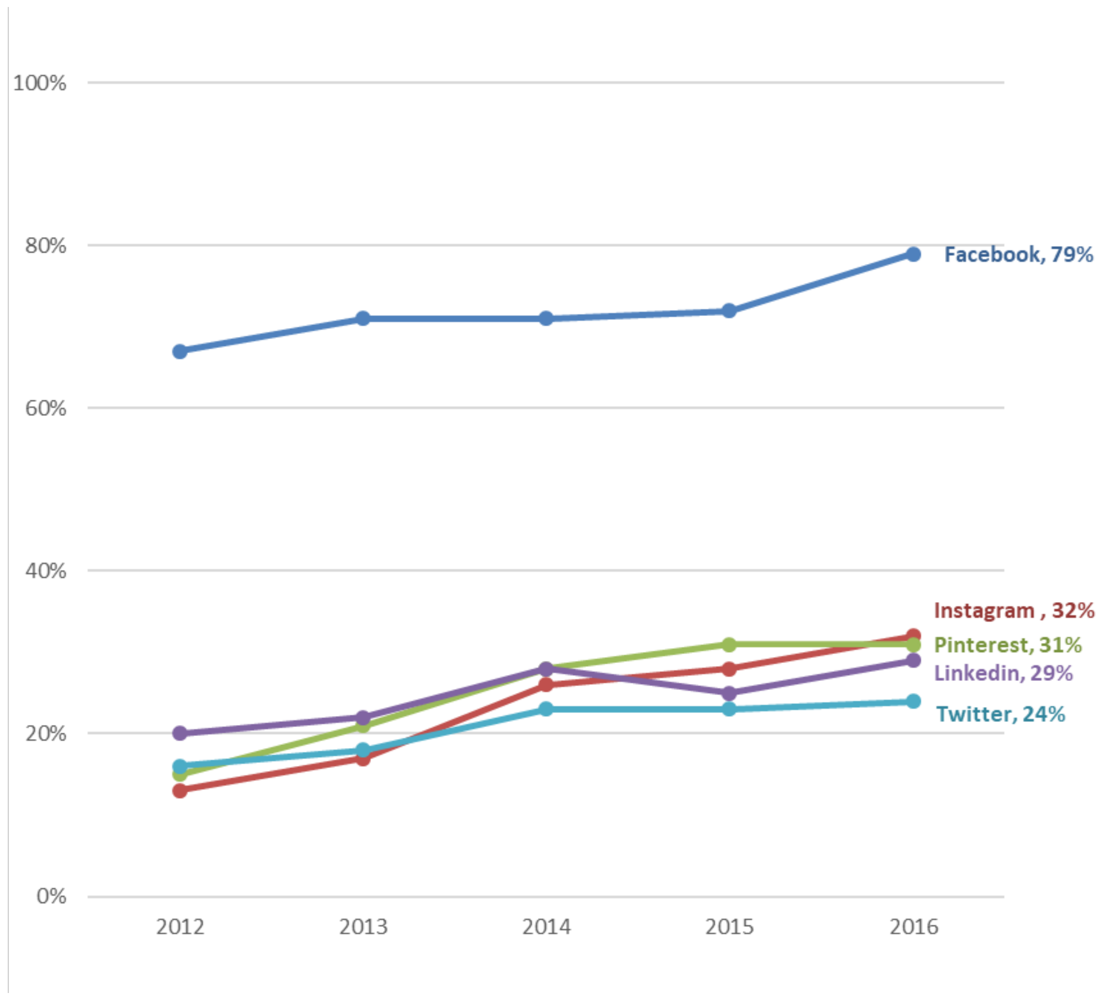

Figure 1: Percentage of Online Adults Who Use Each Social Media Platforms

This figure shows the percentage of online adults who use each social media platform. By April 2016, 86% of Americans are internet users. Source: Pew Research Center, November 2016, "Social Media Update 2016"

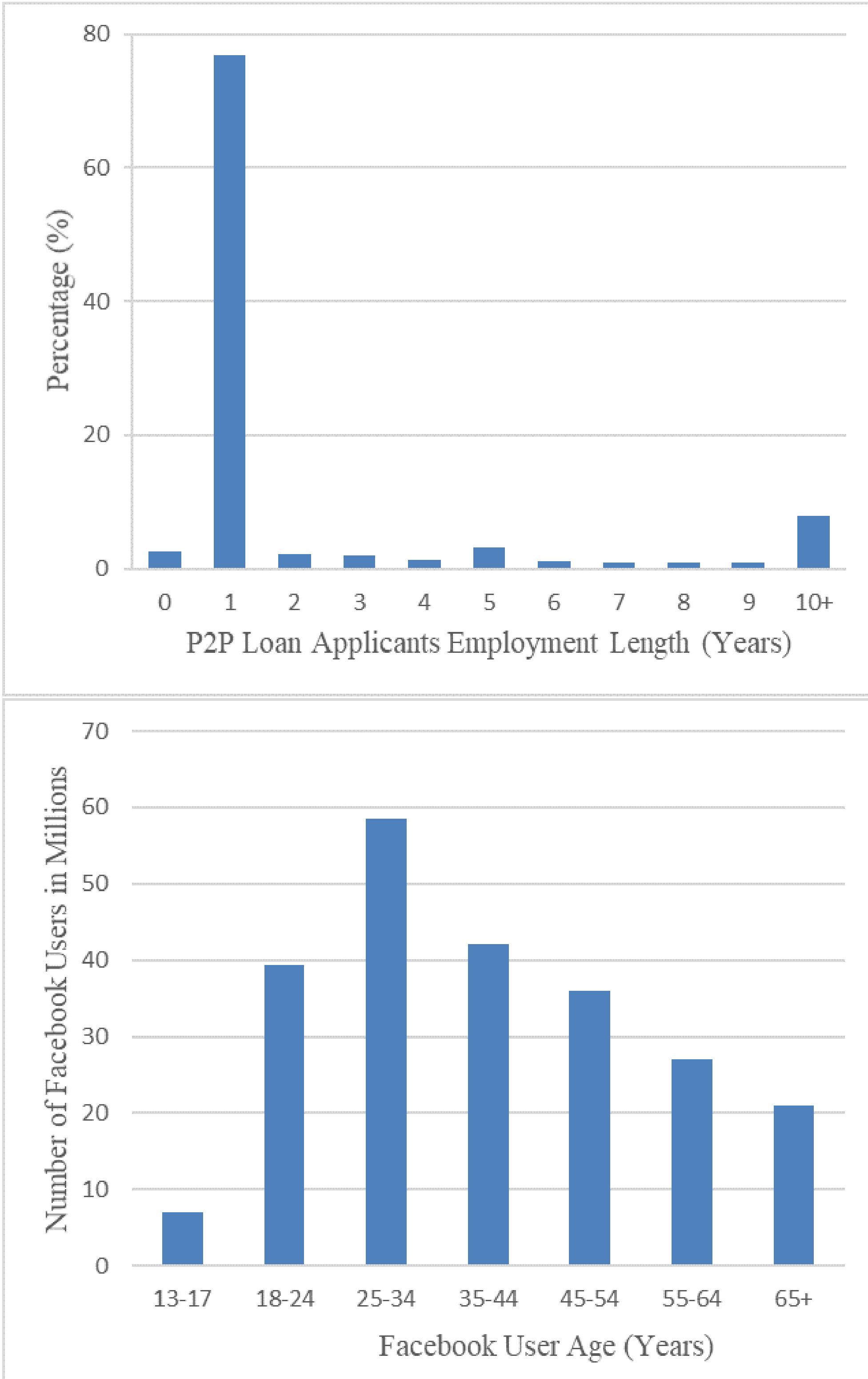

Figure 2: Distribution of LendingClub Borrower Employment Length and Facebook User Age

This figure shows the distribution of LendingClub loan applicants' employment length (by year) and Facebook user age as of 2016.

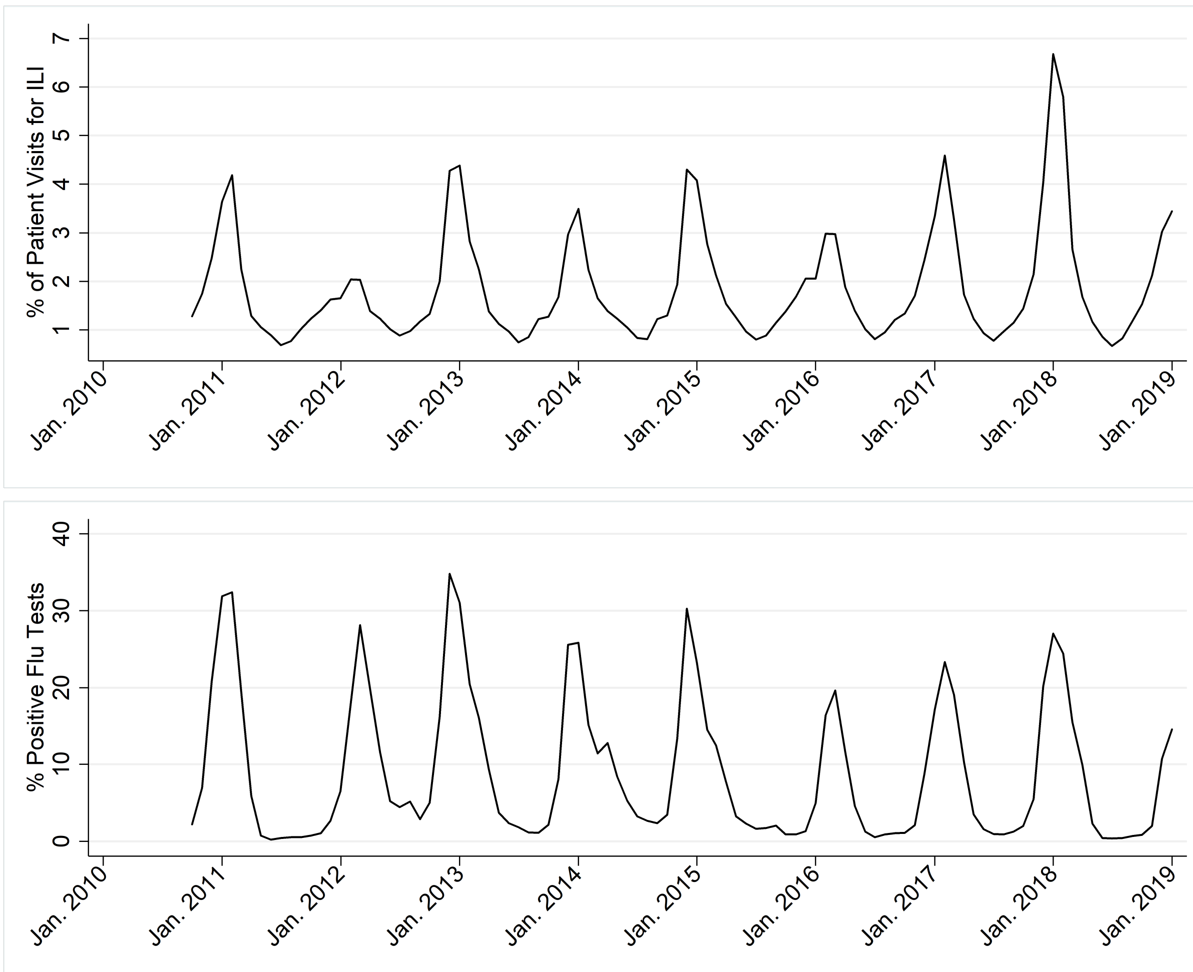

Figure 3: Time Series of Flu Measures

This figure presents the nation-level monthly time series of the two measures of flu epidemics.

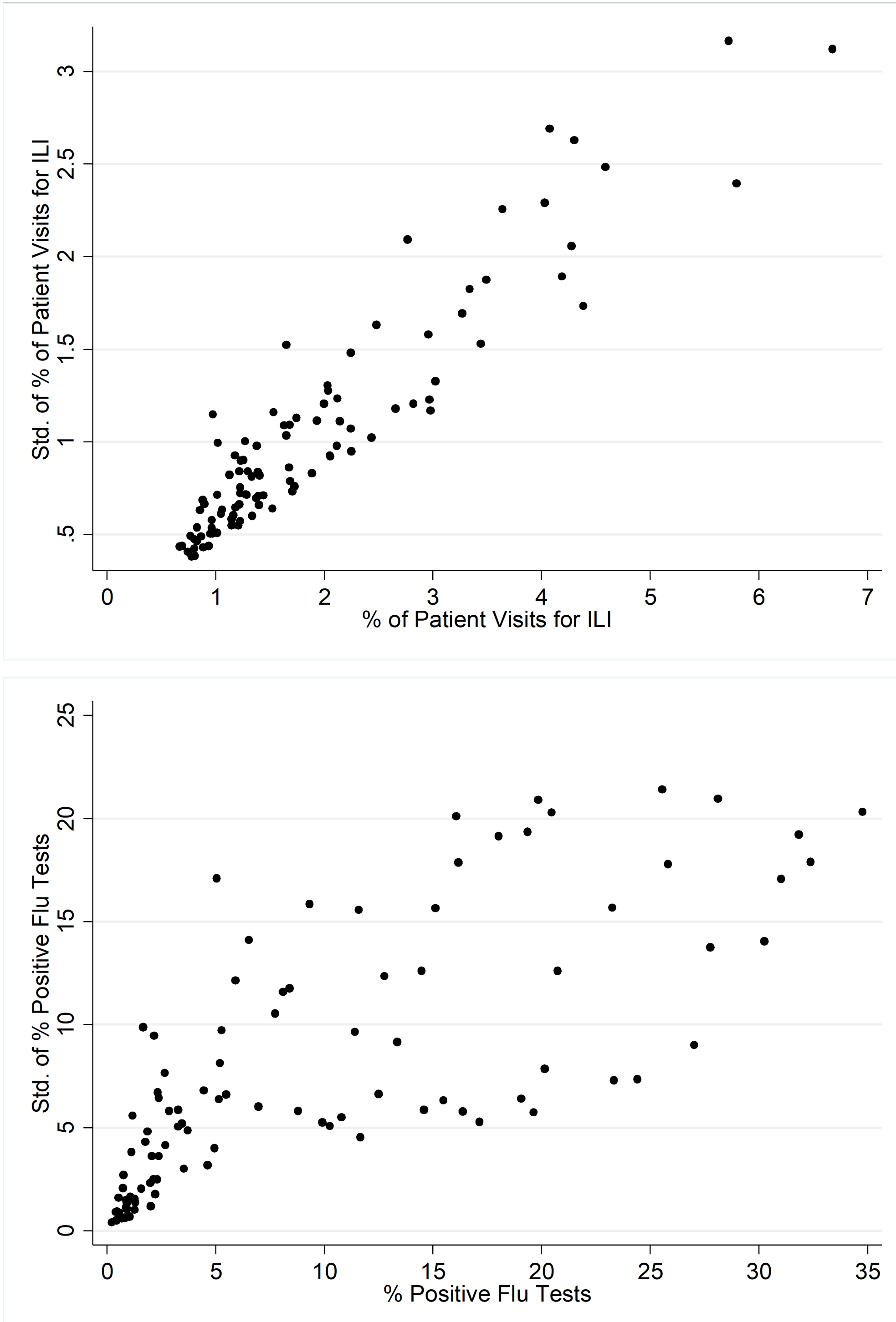

Figure 4: Relationship Between the Level of and Cross-sectional Variation in Flu Measures

This figure shows the relationship between monthly nation-level value of flu measures and monthly cross-sectional variation in flu measures across all U.S. states.

Table 1: Variable Definitions

Variable Name	Definition
SCI_i	$SCI_i = \Sigma_j SCI_{i,j}, j \neq i$ where $SCI_{i,j}$ is the normalized total number of friendship links between area i and area j. $SCI_{i,j}$ has a minimum value of 0 and a maximum value of 1000000. The relative differences in the $SCI_{i,j}$ correspond to relative differences in the total number of friendship links. SCI_i is thus area i's total normalized number of friends in all other areas.
SCI_i^{OS}	$SCI_i^{OS} = \Sigma_j SCI_{i,j}^{OS}$, $state(j) \neq state(i)$ $SCI_{i,j}^{OS}$ has the same definition with $SCI_{i,j}$ except that area i and j have to be in different states. Thus, SCI_i^{OS} measures area i's total normalized number of friends in all other states.
$RSCI_i$	$RSCI_i = \Sigma_j SCI_i / Pop_i$, where Pop_i is the population in county i.
$RSCI_i^{OS}$	$RSCI_i^{OS} = \Sigma_j SCI_i^{OS} / Pop_i$, where Pop_i is the population in county i.
Interest Rate	Interest rate on the loan.
Grade	LC assigned loan grade.
FundAmnt	The total amount committed to that loan at that point in time.
Employment Length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
FICO	The simple average of the upper and lower boundary ranges the borrower's FICO at loan origination belongs to.
Income	The self-reported annual income provided by the borrower during registration.
DTI	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
Home Ownership	The home ownership status provided by the borrower during registration. Values are: RENT, OWN, MORTGAGE, OTHER.
Population	The total population of all counties containing a specific 3-digit zip code.
Branch	Number of bank branches per one thousand population. Source: FDIC Summary of Deposits.
Deposit	Amount of bank deposit (in thousands) per capita. Source: FDIC Summary of Deposits.
HHI	Herfindahl-Hirschmann index computed in terms of deposits. Source: FDIC Summary of Deposits
White	Share of white people in local population.
Female	Share of female in local population.
Education	The number of people over 25 years and with high school education (or higher) divided by total population older than 25. Source: 2013 American Community Survey 5-year estimates
Income	Income per capita. Source: 2013 American Community Survey 5-year estimates
Unemployment	Local unemployment rate. Source: 2013 American Community Survey 5-year estimates
Senior	The share of population that is over 65 years old. Source: 2013 American Community Survey 5-year estimates
Manufacture	The share of labor working in manufacturing industry. Source: 2013 American Community Survey 5-year estimates
Infomration	The share of labor working in information industry. Source: 2013 American Community Survey 5-year estimates
ILI	The percentage of patient visits to healthcare provider for influenza-like illness. Source: CDC and WHO/NREVSS
POS	The percentage of flu tests with positive results. Source: CDC and WHO/NREVSS

Table 2: Summary Statistics

This table presents summary statistics for our sample that spans the period of Jan 2010 to Dec 2017. Panel A shows summary statistics for the full loan application sample and origination sample, including both approved and rejected loans. Panel B presents summary statistics for approved loans only. Panel C and Panel D presents the distribution of monthly loan status for all loans originated in 2015. We focus on monthly loan statuses for each of the 24 months following loan origination. For each loan in each of the 24 months following loan origination, the status could be one of the followings: *Issued*, *Current*, *In Grace Period*, *Late (16-30 days)*, *Late (31-120 days)*, *Default*, *Charged Off*, and *Fully Paid*. A loan is at the status of *Issued* in the origination month. The status is *Current* if a monthly payment is made before due date. If a borrower misses a monthly payment and the end of that month is within 1-15 days past the due date, then the status becomes *In Grace Period* for that month. If a payment is not received within 16-30 days after the due date, the status becomes *Late (16-30 days)*. If a payment was not received within 31-120 days after the due date, the status becomes *Late (31-120 days)*. When borrowers miss several payments, the loan will enter *Default* status and, when there is no longer a reasonable expectation of further borrower payments, the loan will be *Charged Off*. The status becomes *Fully Paid* if the loan is fully paid off. A loan is defined as a normal loan if its status is *Issued*, *Current*, *Fully Paid*, or *In Grace Period*. A loan is defined as an abnormal loan if its status is *Late (16-30 days)*, *Late (31-120 days)*, *Default*, or *Charged Off*. Panel D shows the number of normal and abnormal loans originated in 2015.

Panel A	Application Sample			Origination Sample		
Variable	Observations	Mean	Median	Observations	Mean	Median
Social Connectivity						
$RSCI$	4,097,363	0.132	0.125	1,489,527	0.133	0.126
$RSCI^{OS}$	4,097,363	0.075	0.069	1,489,527	0.075	0.070
Loan and Borrower Charateristics						
Approval	4,097,363	0.363	0	N.A.	N.A.	N.A.
Loan Amnt ($Thousands)	4,097,363	13.142	10.000	1,489,527	14.761	12.55
ln(Loan Amnt)	4,097,363	2.169	2.303	1,489,527	2.483	2.53
FICO	4,097,363	699.817	690	1,489,527	698.047	692
DTI	4,097,363	29.548	20.02	1,489,527	18.667	17.96
Employment Length	4,097,363	2.145	1.000	1,489,527	5.697	6
Interest Rate	N.A.	N.A.	N.A.	1,489,527	13.206	12.74
Loan Grade	N.A.	N.A.	N.A.	1,489,527	24.357	25
Demographic Characteristics						
Population	4,097,363	1.063	0.613	1,489,527	1.100	0.633
White	4,097,363	0.738	0.768	1,489,527	0.740	0.768
Female	4,097,363	0.507	0.508	1,489,527	0.507	0.508
Senior	4,097,363	0.136	0.130	1,489,527	0.135	0.129
Economic Characteristics						
Income	4,097,363	27.901	26.813	1,489,527	28.855	27.726
Unemployment	4,097,363	0.099	0.099	1,489,527	0.098	0.098
Manufacture	4,097,363	0.789	0.143	1,489,527	0.732	0.132
Information	4,097,363	0.128	0.033	1,489,527	0.126	0.035
Branch	4,097,363	0.284	0.276	1,489,527	0.285	0.279
Deposit	4,097,363	29.003	19.283	1,489,527	30.308	19.944
HHI	4,097,363	0.148	0.123	1,489,527	0.148	0.122
Social Characteristics						
Education	4,097,363	0.857	0.868	1,489,527	0.861	0.872
Flu Epidemic Measures						
ILI	88,678	0.019	0.014	88,678	0.019	0.014
POS	88,678	0.092	0.050	88,678	0.092	0.050

Panel B			
Monthly Status	Number of Obs.	Percent	Cumulative Percent
Issued	299	0.00%	0.00%
Current	7,924,917	94.96%	94.96%
Fully Paid	134,873	1.62%	96.58%
In Grace Period	139	0.00%	96.58%
Late (16-30 days)	58,798	0.70%	97.28%
Late (31-120 days)	171,991	2.06%	99.34%
Default	7,680	0.09%	99.43%
Charged off	47,009	0.57%	100.00%

Panel C	
Loan Status	Number of Loans
Normal Loans	354,803
Abnormal Loans	66,292
Total Loans	421,095

Table 3: Correlation Matrix

This table reports the correlation matrix for the variables in the origination sample.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1)	$Approved$	1.000																		
(2)	$RSCI$	0.017	1.000																	
(3)	$RSCI^{OS}$	0.018	0.794	1.000																
(4)	$LoanAmount$	-0.001	-0.012	-0.013	1.000															
(5)	$FICO$	0.003	-0.001	-0.004	0.084	1.000														
(6)	DTI	-0.233	-0.017	-0.024	0.071	0.015	1.000													
(7)	$EmploymentLength$	0.654	0.005	0.007	0.038	0.037	-0.231	1.000												
(8)	$Population$	0.018	-0.103	0.000	-0.006	0.004	-0.036	0.020	1.000											
(9)	$Branches$	-0.008	0.107	0.114	0.012	0.007	0.014	0.001	-0.260	1.000										
(10)	$Deposit$	0.031	0.249	0.307	-0.003	0.013	-0.033	0.014	0.263	0.152	1.000									
(11)	HHI	0.008	0.119	0.191	-0.015	-0.003	-0.006	-0.003	0.050	-0.170	0.608	1.000								
(12)	$\%White$	-0.017	-0.381	-0.386	0.010	0.002	0.041	-0.016	-0.388	0.358	-0.234	-0.173	1.000							
(15)	$\%Female$	0.008	0.229	0.201	-0.019	-0.003	-0.025	0.005	0.165	0.006	0.124	0.077	-0.388	1.000						
(16)	$Education$	0.042	0.191	0.219	0.020	0.006	-0.023	0.028	-0.247	0.281	0.061	0.029	0.297	-0.053	1.000					
(17)	$\%Income$	0.074	0.207	0.256	0.039	0.023	-0.069	0.062	0.100	0.116	0.354	0.077	-0.097	0.046	0.581	1.000				
(18)	$Unemployment$	-0.038	-0.231	-0.159	0.003	-0.011	-0.012	0.021	0.027	-0.155	-0.177	-0.032	-0.096	-0.037	-0.444	-0.321	1.000			
(19)	$\%Senior$	-0.033	-0.290	-0.194	-0.007	0.000	0.024	-0.029	-0.251	0.226	-0.186	-0.125	0.383	-0.077	0.069	-0.153	0.101	1.000		
(20)	$\%Manufacuture$	-0.013	-0.019	-0.037	0.001	-0.003	0.023	-0.014	-0.105	0.094	-0.164	-0.172	0.135	-0.018	-0.017	-0.172	-0.024	0.007	1.000	
(21)	$\%Information$	0.001	0.052	0.055	0.009	0.000	0.011	-0.003	-0.043	0.048	-0.143	-0.158	0.056	-0.003	0.087	-0.031	-0.062	-0.094	0.827	1.000

Table 4: Social Connectivity and P2P Borrowing Decisions

This table shows the results of first stage of the 2SLS regression:

$$FrndExp^{All}_{i,t_1,t_2} = \beta^{FS} FrndExp^{OS}_{i,t_1,t_2} + \delta \mathbf{X}_{i,t_1} + \epsilon_{i,t_1,t_2}$$

where $FrndExp^{All}_{i,t_1,t_2}$ is the average borrowing experience of area i's all connections between time t_1 and t_2. $FrndExp^{OS}_{i,t_1,t_2}$ is the average borrowing experience of area i's out-of-state connections between time t_1 and t_2. Friend borrowing experience between t_1 and t_2 includes changes in loan application ($\Delta FrndApp_{i,t_1,t_2}$), changes in loan origination ($\Delta FrndOrg_{i,t_1,t_2}$), changes in approval rate ($\Delta FrndRate_{i,t_1,t_2}$), and aggregate approval rate ($FrndRate_{i,0}$). Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$\Delta FrndOrg^{All}_{i,t_1,t_2}$	$\Delta FrndApp^{All}_{i,t_1,t_2}$	$\Delta FrndRate^{All}_{i,t_1,t_2}$	$FrndRate^{All}_{i,t_1,t_2}$
	(1)	(2)	(3)	(4)
$\Delta FrndOrg^{OS}_{i,t_1,t_2}$	0.265***			
	(0.062)			
$\Delta FrndApp^{OS}_{i,t_1,t_2}$		0.205***		
		(0.060)		
$\Delta FrndRate^{OS}_{i,t_1,t_2}$			29.192***	
			(8.227)	
$FrndRate^{OS}_{i,t_1,t_2}$				26.707***
				(6.365)
$Branch_{i,t}$	-0.015**	-0.093**	-0.058	-0.424**
	(0.007)	(0.037)	(0.101)	(0.193)
$Deposit_{i,t}$	0.063*	0.389*	0.679	1.839*
	(0.033)	(0.194)	(0.507)	(0.971)
$HHI_{i,t}$	-0.008	-0.031	-0.009	-0.163
	(0.005)	(0.031)	(0.077)	(0.164)
$Population_{i,t}$	-0.866***	-1.783	-16.576***	-31.472***
	(0.300)	(1.588)	(3.611)	(8.155)
$White_{i,t}$	0.005	0.049	0.178	0.262
	(0.007)	(0.045)	(0.121)	(0.249)
$Female_{i,t}$	0.135	0.668	1.015	4.186
	(0.089)	(0.500)	(1.279)	(3.024)
$Education_{i,t}$	0.017	0.090	-0.082	0.402
	(0.022)	(0.129)	(0.315)	(0.741)
$Income_{i,t}$	-0.001***	-0.003**	-0.007**	-0.014**
	0.000	(0.001)	(0.003)	(0.007)
$Unemployment_{i,t}$	-0.008	-0.045	0.373	0.920
	(0.062)	(0.291)	(0.753)	(1.796)
$Senior_{i,t}$	-0.024	-0.150	-0.242	-0.931
	(0.027)	(0.159)	(0.369)	(0.807)
$Manufacture_{i,t}$	-0.001	-0.010	-0.021*	-0.043
	(0.001)	(0.006)	(0.011)	(0.028)
$Information_{i,t}$	0.010**	0.058**	0.149**	0.296**
	(0.004)	(0.026)	(0.061)	(0.127)
Observations	854	854	854	854
Adjusted R^2	0.985	0.985	0.981	0.983
State FE	Yes	Yes	Yes	Yes

Table 5: Social Connectivity and P2P Borrowing Decisions

This table reports the results of the second stage of the 2SLS regressions of borrowing decisions and outcomes on friends' past experience on P2P lending: $(abcdefg)\widehat{}$

$$Application_{i,t_2,t_3} = \beta^{IV}\widehat{FrndExp^{All}_{i,t_1,t_2}} + \gamma\mathbf{X}_{i,t_2} + \epsilon_{i,t_2,t_3}.$$

where i indexes area. t_1, t_2, t_3 index the year 2010, 2013, and 2016, respectively. $Application_{i,t_2,t_3}$ is the percentage change in loan application between t_2 and t_3. $\widehat{FrndExp^{All}_{i,t_1,t_2}}$ is the estimated average borrowing experience of area i's all connections between time t_1 and t_2, including changes in count (or amount) of loan application ($\Delta FrndApp_{i,t_1,t_2}$), changes in loan origination ($\Delta FrndOrg_{i,t_1,t_2}$), changes in approval rate ($\Delta FrndRate_{i,t_1,t_2}$), and aggregate approval rate ($FrndRate_{i,t_1,t_2}$). All application and originations on scaled by local population. Areas are at the 3-digit zip level. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$\Delta Application_{i,t_2,t_3}$			
	(1)	(2)	(3)	(4)
$\Delta FrndOrg_{i,t_1,t_2}$	11.823***			
	(4.392)			
$\Delta FrndApp_{i,t_1,t_2}$		2.535**		
		(1.118)		
$\Delta FrndRate_{i,t_1,t_2}$			79.309**	
			(36.456)	
$FrndRate_{i,t_1,t_2}$				34.494**
				(13.833)
$Branch_{i,t}$	-0.888**	-0.837**	-1.052***	-0.925***
	(0.359)	(0.371)	(0.348)	(0.337)
$Deposit_{i,t}$	1.255	0.984	1.683	1.523
	(1.952)	(1.995)	(1.949)	(1.944)
$HHI_{i,t}$	-0.820**	-0.836***	-0.909***	-0.865***
	(0.320)	(0.311)	(0.343)	(0.328)
$Population_{i,t}$	27.747**	21.091*	32.065**	29.134***
	(11.004)	(10.974)	(12.935)	(11.182)
$White_{i,t}$	-0.277	-0.349	-0.406	-0.329
	(0.378)	(0.396)	(0.383)	(0.377)
$Female_{i,t}$	6.015***	5.901**	6.790***	6.047***
	(2.248)	(2.404)	(2.096)	(2.216)
$Education_{i,t}$	0.028	-0.054	0.268	0.131
	(0.573)	(0.610)	(0.516)	(0.558)
$Income_{i,t}$	-0.004	-0.002	-0.004	-0.006
	(0.006)	(0.006)	(0.008)	(0.007)
$Unemployment_{i,t}$	1.458	1.308	0.987	1.142
	(1.335)	(1.420)	(1.471)	(1.334)
$Senior_{i,t}$	0.826	0.915	0.750	0.869
	(0.942)	(1.098)	(0.886)	(0.885)
$Manufacture_{i,t}$	-0.010	-0.001	-0.011	-0.013
	(0.022)	(0.026)	(0.023)	(0.022)
$Information_{i,t}$	0.000	-0.032	-0.006	0.018
	(0.148)	(0.165)	(0.160)	(0.147)
Observations	807	807	807	807
Adjusted R^2	0.760	0.745	0.748	0.760
State FE	Yes	Yes	Yes	Yes

Table 6: Social Connectivity and Loan-level Approval

This table reports the coefficient estimates from the following logit regression:

$$Approval_{i,k} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i indexes area and k indexes loan application. Panel A estimates the regression using the whole sample and Panel B estimates the regression for three separate sub-samples sorted by FICO scores. $Approval_{i,k}$ is 1 for approved loans and 0 for rejected loans. $RSCI_i$ is the relative social connectedness index for county i measuring its total connectivity with all other regions of the country. $\boldsymbol{X}_k$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. s_i indexes state fixed effects and q_k indexes application quarter fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

Panel A	$Approval_{i,k}$		
	(1)	(2)	(3)
$RSCI_i$	2.241***		0.662**
	(0.429)		(0.261)
$ln(Amount)_k$		-0.066***	-0.066***
		(0.012)	(0.012)
$FICO_k$		0.002***	0.002***
		(0.000)	(0.000)
DTI_k		-0.027***	-0.027***
		(0.001)	(0.001)
$EmploymentLength_k$		3.027***	3.027***
		(0.024)	(0.024)
$Population_{i,t}$		-25.347***	-21.414***
		(4.640)	(5.731)
$Branch_{i,t}$		-0.001***	-0.000***
		(0.000)	(0.000)
$Deposit_{i,t}$		1.400***	1.253***
		(0.365)	(0.272)
$HHI_{i,t}$		-0.143*	-0.140*
		(0.085)	(0.072)
$White_{i,t}$		0.003	0.093
		(0.065)	(0.078)
$Female_{i,t}$		1.468	1.753
		(1.368)	(1.463)
$Education_{i,t}$		1.236***	1.099***
		(0.334)	(0.333)
$Income_{i,t}$		0.008***	0.008***
		(0.002)	(0.002)
$Unemployment_{i,t}$		0.675	0.753
		(0.490)	(0.483)
$Senior_{i,t}$		-0.039	0.052
		(0.453)	(0.441)
$Manufacture_{i,t}$		-0.008**	-0.009**
		(0.003)	(0.003)
$Information_{i,t}$		0.093***	0.105***
		(0.026)	(0.027)
Observations	4,097,888	4,097,363	4,097,363
Pseudo R^2	0.135	0.433	0.433
State FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes

Table 7: Social Connectivity and Loan-level Approval

This table reports the coefficient estimates from the following logistic regression for three sub-samples divided by FICO scores:

$$Approval_{i,k} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i indexes area and k indexes loan application. $Approval_{i,k}$ is 1 for approved loans and 0 for rejected loans. $RSCI_i$ is the relative social connectedness index for county i measuring its total connectivity with all other regions of the country. $\boldsymbol{X}_{i,k}$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. s_k indexes state fixed effects and q_k indexes application quarter fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$Approval_{i,k}$		
	(1)	(2)	(3)
	Low FICOs	Medium FICOs	High FICOs
	(660 - 699)	(700 - 749)	(750 -)
$RSCI_i$	0.591**	0.701**	1.031***
	(0.258)	(0.290)	(0.384)
$ln(Amount)_k$	-0.109***	-0.045***	-0.223***
	(0.013)	(0.011)	(0.019)
$FICO_k$	0.003***	0.003***	0.003***
	0.000	0.000	0.000
DTI_k	-0.033***	-0.028***	-0.031***
	(0.001)	(0.001)	(0.001)
$EmploymentLength_k$	3.176***	2.970***	2.736***
	(0.026)	(0.026)	(0.028)
$Population_{i,t}$	-22.135***	-13.461*	-12.598
	(5.100)	(8.111)	(9.154)
$Branch_{i,t}$	-0.001***	-0.000***	-0.000***
	0.000	0.000	0.000
$Deposit_{i,t}$	1.273***	1.860***	0.757*
	(0.227)	(0.323)	(0.458)
$HHI_{i,t}$	-0.124*	-0.255***	-0.025
	(0.070)	(0.073)	(0.147)
$White_{i,t}$	0.060	-0.022	0.323**
	(0.071)	(0.101)	(0.147)
$Female_{i,t}$	2.070	1.307	1.350
	(1.361)	(1.331)	(2.458)
$Education_{i,t}$	0.917***	1.299***	2.286***
	(0.318)	(0.357)	(0.603)
$Income_{i,t}$	0.009***	0.010***	-0.001
	(0.002)	(0.002)	(0.003)
$Unemployment_{i,t}$	0.703	0.772	1.235
	(0.460)	(0.523)	(0.836)
$Senior_{i,t}$	0.052	-0.051	0.421
	(0.427)	(0.487)	(0.622)
$Manufacture_{i,t}$	-0.008**	-0.010**	-0.009
	(0.003)	(0.004)	(0.008)
$Information_{i,t}$	0.106***	0.089***	0.098**
	(0.027)	(0.032)	(0.049)
Observations	2,460,157	1,231,549	359,462
Pseudo R^2	0.468	0.430	0.417
State FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes

Table 8: Social Connectivity and Areal-level Loan Approval Rate: Tobit Regressions

This table reports the coefficient estimates from the following Tobit regression:

$$ApprovalRate_{i,t} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i and t indexes region and quarter. $ApprovalRate_{i,t}$ is the number (or amount) of approved loans in region i in quarter t divided by the total number (or amount) applications. $\boldsymbol{X}_k$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. s_i indexes state fixed effects and q_k indexes application quarter fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$ApprovalRate_{i,t}$					
	Count			Amount		
	(1)	(2)	(3)	(4)	(5)	(6)
$RSCI_i$	24.168***		6.748***	35.304***		7.331**
	(4.909)		(2.495)	(5.123)		(2.996)
$ln(Amount)_{i,t}$		0.830**	0.912**		-5.755***	-5.664***
		(0.389)	(0.389)		(0.400)	(0.397)
$FICO_{i,t}$		0.032***	0.032***		0.036***	0.036***
		(0.002)	(0.002)		(0.002)	(0.002)
$DTI_{i,t}$		0.001	0.001		0.000	0.000
		(0.002)	(0.002)		(0.002)	(0.002)
$EmploymentLength_{i,t}$		6.657***	6.662***		8.251***	8.257***
		(0.086)	(0.087)		(0.112)	(0.112)
$Population_{i,t}$		0.890***	1.010***		0.948***	1.078***
		(0.208)	(0.224)		(0.201)	(0.217)
$Branch_{i,t}$		-0.003***	-0.003***		-0.005***	-0.005***
		(0.001)	(0.001)		(0.001)	(0.001)
$Deposit_{i,t}$		26.099***	25.364***		28.303***	27.505***
		(6.032)	(5.492)		(7.123)	(6.515)
$HHI_{i,t}$		-3.697***	-3.696***		-4.082***	-4.081***
		(0.903)	(0.838)		(1.100)	(1.019)
$White_{i,t}$		1.515*	2.321***		0.379	1.255
		(0.810)	(0.827)		(0.730)	(0.768)
$Female_{i,t}$		-1.997	-2.060		-3.216	-3.279
		(7.114)	(7.339)		(7.604)	(7.812)
$Education_{i,t}$		2.781	1.613		4.793**	3.526*
		(2.196)	(2.120)		(2.085)	(2.076)
$Income_{i,t}$		0.215***	0.208***		0.247***	0.240***
		(0.024)	(0.027)		(0.021)	(0.024)
$Unemployment_{i,t}$		-2.283	-2.334		3.186	3.129
		(4.647)	(4.775)		(4.912)	(5.016)
$Senior_{i,t}$		-6.368	-4.787		-10.126**	-8.409*
		(4.002)	(3.767)		(4.668)	(4.386)
$Manufacture_{i,t}$		-0.004	-0.002		-0.035	-0.032
		(0.035)	(0.038)		(0.039)	(0.043)
$Information_{i,t}$		0.422	0.456		0.501	0.538
		(0.313)	(0.326)		(0.353)	(0.367)
Observations	55,988	55,988	55,988	55,988	55,988	55,988
Pseudo R^2	0.208	0.259	0.259	0.186	0.251	0.251
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: Social Connectivity and Areal-level Loan Approval Rate: Linear Regressions

This table reports the coefficient estimates from the following OLS regression:

$$ApprovalRate_{i,t} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i and t indexes region and quarter. Observations with approval rates being 0% or 100%, which constitute 5% of the sample size, are dropped out of the sample to allow linear regressions. $ApprovalRate_{i,t}$ is the number (or amount) of approved applications in region i in quarter t divided by the total number (or amount) of applications. $\boldsymbol{X}_k$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. s_i indexes state fixed effects and q_k indexes application quarter fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$ApprovalRate_{i,t}$					
	Count			Amount		
	(1)	(2)	(3)	(4)	(5)	(6)
$RSCI_i$	21.515***		7.503***	32.928***		7.991**
	(4.542)		(2.570)	(4.685)		(3.053)
$ln(Amount)_{i,t}$		1.094***	1.191***		-6.274***	-6.171***
		(0.377)	(0.377)		(0.376)	(0.376)
$FICO_{i,t}$		0.029***	0.029***		0.034***	0.034***
		(0.002)	(0.002)		(0.002)	(0.002)
$DTI_{i,t}$		0.000	0.000		-0.001	-0.001
		(0.001)	(0.001)		(0.002)	(0.002)
$EmploymentLength_{i,t}$		6.543***	6.549***		8.233***	8.239***
		(0.082)	(0.082)		(0.103)	(0.103)
$Population_{i,t}$		0.329*	0.461**		0.422**	0.562***
		(0.173)	(0.205)		(0.160)	(0.189)
$Branch_{i,t}$		-0.002**	-0.002**		-0.003***	-0.003***
		(0.001)	(0.001)		(0.001)	(0.001)
$Deposit_{i,t}$		22.159***	21.350***		24.025***	23.163***
		(5.798)	(5.213)		(6.738)	(6.088)
$HHI_{i,t}$		-2.484***	-2.486***		-2.850***	-2.852***
		(0.823)	(0.759)		(0.972)	(0.889)
$White_{i,t}$		1.167	2.069***		0.104	1.064
		(0.734)	(0.683)		(0.651)	(0.640)
$Female_{i,t}$		-5.046	-5.053		-6.730	-6.737
		(6.012)	(6.313)		(6.515)	(6.794)
$Education_{i,t}$		-0.021	-1.333		1.849	0.452
		(2.000)	(1.799)		(1.767)	(1.587)
$Income_{i,t}$		0.201***	0.193***		0.239***	0.231***
		(0.024)	(0.028)		(0.020)	(0.024)
$Unemployment_{i,t}$		-7.441*	-7.488*		-0.994	-1.043
		(4.132)	(4.168)		(4.597)	(4.584)
$Senior_{i,t}$		-5.040	-3.300		-8.764**	-6.911*
		(3.504)	(3.301)		(4.152)	(3.879)
$Manufacture_{i,t}$		-0.021	-0.019		-0.054	-0.052
		(0.034)	(0.037)		(0.038)	(0.041)
$Information_{i,t}$		0.482	0.522		0.583*	0.626*
		(0.300)	(0.313)		(0.338)	(0.351)
Observations	53,285	53,285	53,285	53,285	53,285	53,285
Adjusted R^2	0.855	0.908	0.908	0.818	0.898	0.898
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 10: Flu Epidemics as Exogenous Shocks to Social Connectivity

This table reports the coefficient estimates from the following panel regression:

$$ApprovalRate_{i,t} = \beta_1 RSCI_i^{OS} + \beta_2 FluImpact_{i,t} + \gamma X_{i,t} + s_k + \theta_t + \epsilon_{i,t}$$

where i and t denote area and time. $ApprovalRate_{i,t}$ is the loan approval rate (total origination count or amount divided by total application count or amount) in area i and month t. $X_{i,t}$ is a vector of loan, borrower, demographic, economic, and social characteristics. s_k indexes state fixed effects and θ_t indexes application time fixed effects. $FluImpact$ is the flu impact measure constructed in Equation (10) and can be either $FluImpact_{i,t}^{ILI}$ or $FluImpact_{i,t}^{POS}$ depending on the flu measure used. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

Panel A	$ApprovalRate_{i,t}$			
	Count	Amount	Count	Amount
$RSCI_i^{OS}$	46.530***	50.762***	20.373***	23.620***
	(3.723)	(3.901)	(2.918)	(3.080)
$FluImpact_{i,t}^{ILI}$	-22.371***	-23.362***		
	(1.441)	(1.348)		
$FluImpact_{i,t}^{POS}$			-1.732***	-1.835***
			(0.160)	(0.158)
$ln(Amount)_{i,t}$	3.002***	-5.515***	3.014***	-5.505***
	(0.405)	(0.449)	(0.408)	(0.449)
$FICO_{i,t}$	0.041***	0.049***	0.041***	0.049***
	(0.003)	(0.004)	(0.003)	(0.004)
$DTI_{i,t}$	-0.066***	-0.077***	-0.064***	-0.075***
	(0.011)	(0.011)	(0.011)	(0.011)
$EmploymentLength_{i,t}$	6.461***	8.453***	6.590***	8.587***
	(0.156)	(0.123)	(0.160)	(0.127)
$Population_{i,t}$	0.827***	0.833***	0.876***	0.884***
	(0.213)	(0.186)	(0.211)	(0.180)
$Branch_{i,t}$	-0.001	-0.002**	0.000	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
$Deposit_{i,t}$	13.826***	17.377***	13.908***	17.457***
	(4.562)	(4.764)	(4.686)	(4.932)
$HHI_{i,t}$	-1.397	-2.066**	-1.324	-1.988**
	(0.868)	(0.948)	(0.858)	(0.937)
$White_{i,t}$	1.000	-0.034	1.663**	0.656
	(0.736)	(0.700)	(0.706)	(0.657)
$Female_{i,t}$	-9.535	-8.663	-9.081	-8.177
	(6.251)	(5.875)	(6.492)	(6.110)
$Education_{i,t}$	-0.314	-0.451	0.498	0.391
	(2.103)	(1.656)	(1.986)	(1.521)
$Income_{i,t}$	0.248***	0.268***	0.236***	0.255***
	(0.026)	(0.023)	(0.023)	(0.021)
$Unemployment_{i,t}$	-4.697	-4.862	-4.836	-4.994
	(5.272)	(4.916)	(5.026)	(4.685)
$Senior_{i,t}$	-3.278	-6.525**	-2.735	-5.959*
	(2.906)	(3.106)	(3.079)	(3.378)
$Manufacture_{i,t}$	-0.048	-0.085**	-0.052	-0.089**
	(0.033)	(0.033)	(0.034)	(0.034)
$Information_{i,t}$	0.614*	0.793***	0.658*	0.840***
	(0.318)	(0.289)	(0.329)	(0.297)
$Flu_{i,t}^{ILI}$	-133.028	-88.964		
	(113.615)	(118.083)		
$Flu_{i,t}^{POS}$			0.008	0.002
			(0.005)	(0.005)
Observations	67,605	67,605	67,605	67,605
Adjusted R^2	0.879	0.885	0.877	0.884
State FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes

Panel B	$ApprovalRate_{i,t}$					
	Count	Amount	Count	Amount	Count	Amount
$RSCI_{i}^{OS}$	12.948***	15.347***	14.414***	17.169***	12.190***	14.740***
	(2.751)	(2.851)	(2.861)	(2.908)	(2.951)	(2.919)
$FluPeak_{i,t}^{ILI}$	-32.166***	-31.563***				
	(3.484)	(4.000)				
$FluPeak_{i,t}^{POS}$			-61.205***	-63.596***		
			(5.864)	(5.379)		
$FluPeak_{i,t}^{ILI,POS}$					-98.636***	-99.775***
					(9.630)	(9.171)
$ln(Amount)_{i,t}$	3.145***	-5.363***	3.083***	-5.430***	3.115***	-5.395***
	(0.408)	(0.450)	(0.410)	(0.451)	(0.410)	(0.452)
$FICO_{i,t}$	0.041***	0.049***	0.041***	0.049***	0.041***	0.049***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
$DTI_{i,t}$	-0.068***	-0.080***	-0.066***	-0.077***	-0.068***	-0.079***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$EmploymentLength_{i,t}$	6.634***	8.634***	6.617***	8.616***	6.617***	8.617***
	(0.161)	(0.128)	(0.161)	(0.127)	(0.160)	(0.127)
$Population_{i,t}$	0.858***	0.865***	0.881***	0.889***	0.868***	0.876***
	(0.211)	(0.183)	(0.211)	(0.181)	(0.212)	(0.183)
$Branch_{i,t}$	0.000	-0.002**	0.000	-0.002**	0.000	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$Deposit_{i,t}$	14.390***	17.970***	14.330***	17.907***	14.587***	18.171***
	(4.661)	(4.909)	(4.734)	(4.980)	(4.696)	(4.945)
$HHI_{i,t}$	-1.455*	-2.128**	-1.407	-2.076**	-1.470*	-2.141**
	(0.858)	(0.944)	(0.869)	(0.945)	(0.865)	(0.947)
$White_{i,t}$	1.459**	0.453	1.628**	0.618	1.499**	0.490
	(0.709)	(0.667)	(0.707)	(0.660)	(0.704)	(0.660)
$Female_{i,t}$	-8.471	-7.541	-8.731	-7.799	-8.509	-7.564
	(6.431)	(6.073)	(6.495)	(6.120)	(6.455)	(6.080)
$Education_{i,t}$	0.178	0.076	0.582	0.477	0.500	0.389
	(1.988)	(1.547)	(1.989)	(1.502)	(1.951)	(1.502)
$Income_{i,t}$	0.238***	0.257***	0.236***	0.255***	0.235***	0.254***
	(0.023)	(0.021)	(0.022)	(0.020)	(0.022)	(0.020)
$Unemployment_{i,t}$	-4.316	-4.454	-4.477	-4.608	-4.483	-4.620
	(4.975)	(4.620)	(5.069)	(4.705)	(5.072)	(4.718)
$Senior_{i,t}$	-2.662	-5.877*	-2.636	-5.853*	-2.725	-5.937*
	(3.012)	(3.273)	(3.089)	(3.383)	(3.000)	(3.274)
$Manufacture_{i,t}$	-0.053	-0.089**	-0.052	-0.089**	-0.053	-0.090***
	(0.034)	(0.034)	(0.034)	(0.033)	(0.034)	(0.033)
$Information_{i,t}$	0.643*	0.824***	0.653*	0.835***	0.650**	0.831***
	(0.321)	(0.292)	(0.326)	(0.294)	(0.320)	(0.291)
$Flu_{i,t}^{ILI}$	-144.839	-101.504			-157.756	-110.451
	(109.415)	(114.859)			(107.260)	(113.028)
$Flu_{i,t}^{POS}$			0.009*	0.003	0.009*	0.003
			(0.005)	(0.005)	(0.005)	(0.005)
Observations	67,605	67,605	67,605	67,605	67,605	67,605
Pseudo R^2	0.876	0.883	0.877	0.883	0.877	0.883
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 11: Flu Epidemics as Exogenous Shocks to Social Connectivity: Tobit Regressions

This table reports the coefficient estimates from the following Tobit regression:

$$ApprovalRate_{i,t} = \beta_1 RSCI_i^{OS} + \beta_2 FluImpact_{i,t} + \gamma X_{i,t} + s_k + \theta_t + \epsilon_{i,t}$$

where i and t denote area and time. $ApprovalRate_{i,t}$ is the loan approval rate (total origination count or amount divided by total application count or amount) in area i and month t. $X_{i,t}$ is a vector of loan, borrower, demographic, economic, and social characteristics. $FluImpact$ is the flu impact measure constructed in Equation (10) and can be either $FluImpact_{i,t}^{ILI}$ or $FluImpact_{i,t}^{POS}$ depending on the flu measure used. s_k indexes state fixed effects and $_t$ indexes application time fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

Panel A	$ApprovalRate_{i,t}$			
	Count	Amount	Count	Amount
$RSCI_i^{OS}$	46.844***	52.442***	19.563***	23.652***
	(3.943)	(4.019)	(3.079)	(3.168)
$FluImpact_{i,t}^{ILI}$	-22.906***	-24.200***		
	(1.488)	(1.407)		
$FluImpact_{i,t}^{POS}$			-1.692***	-1.797***
			(0.170)	(0.170)
$ln(Amount)_{i,t}$	2.243***	-4.887***	2.249***	-4.884***
	(0.474)	(0.527)	(0.477)	(0.531)
$FICO_{i,t}$	0.055***	0.066***	0.055***	0.067***
	(0.004)	(0.004)	(0.004)	(0.004)
$DTI_{i,t}$	-0.020	-0.031**	-0.018	-0.029**
	(0.012)	(0.014)	(0.012)	(0.014)
$EmploymentLength_{i,t}$	6.479***	8.176***	6.592***	8.297***
	(0.111)	(0.107)	(0.112)	(0.109)
$Population_{i,t}$	1.381***	1.405***	1.435***	1.462***
	(0.287)	(0.275)	(0.288)	(0.275)
$Branch_{i,t}$	-0.003***	-0.004***	-0.002***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
$Deposit_{i,t}$	21.270***	26.271***	21.415***	26.424***
	(5.131)	(5.434)	(5.159)	(5.482)
$HHI_{i,t}$	-4.130***	-5.248***	-4.090***	-5.203***
	(1.166)	(1.241)	(1.129)	(1.192)
$White_{i,t}$	2.123**	1.009	2.828***	1.749*
	(1.031)	(1.030)	(0.972)	(0.955)
$Female_{i,t}$	-1.685	2.391	-1.087	3.019
	(8.648)	(9.133)	(8.897)	(9.403)
$Education_{i,t}$	6.621**	7.451***	7.456**	8.326***
	(3.066)	(2.409)	(3.010)	(2.356)
$Income_{i,t}$	0.276***	0.285***	0.264***	0.272***
	(0.024)	(0.021)	(0.022)	(0.020)
$Unemployment_{i,t}$	15.541**	15.469**	15.443**	15.384**
	(7.311)	(6.833)	(7.363)	(6.885)
$Senior_{i,t}$	-9.002**	-12.875***	-8.593**	-12.436**
	(3.981)	(4.596)	(4.287)	(4.981)
$Manufacture_{i,t}$	-0.005	-0.034	-0.008	-0.038
	(0.049)	(0.052)	(0.049)	(0.052)
$Information_{i,t}$	0.528	0.614	0.563	0.653*
	(0.375)	(0.388)	(0.381)	(0.394)
$Flu_{i,t}^{ILI}$	-199.920	-155.479		
	(140.567)	(155.336)		
$Flu_{i,t}^{POS}$			0.008	0.003
			(0.008)	(0.009)
Observations	70,991	70,991	70,991	70,991
Pseudo R^2	0.213	0.217	0.211	0.216
State FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes

Panel B	$ApprovalRate_{i,t}$					
	Count	Amount	Count	Amount	Count	Amount
$RSCI_i^{OS}$	12.653***	15.772***	13.918***	17.474***	11.895***	15.307***
	(2.829)	(2.934)	(2.858)	(2.912)	(2.910)	(2.824)
$FluPeak_{i,t}^{ILI}$	-33.966***	-32.812***				
	(4.317)	(4.778)				
$FluPeak_{i,t}^{POS}$			-61.649***	-63.894***		
			(6.112)	(5.620)		
$FluPeak_{i,t}^{ILI,POS}$					-105.119***	-107.250***
					(10.448)	(9.526)
$ln(Amount)_{i,t}$	2.359***	-4.765***	2.303***	-4.826***	2.326***	-4.801***
	(0.475)	(0.530)	(0.478)	(0.533)	(0.478)	(0.534)
$FICO_{i,t}$	0.056***	0.067***	0.055***	0.067***	0.055***	0.067***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$DTI_{i,t}$	-0.022*	-0.033**	-0.020	-0.031**	-0.021*	-0.032**
	(0.012)	(0.014)	(0.012)	(0.014)	(0.012)	(0.014)
$EmploymentLength_{i,t}$	6.630***	8.338***	6.614***	8.321***	6.614***	8.322***
	(0.111)	(0.108)	(0.112)	(0.109)	(0.112)	(0.108)
$Population_{i,t}$	1.419***	1.446***	1.441***	1.468***	1.428***	1.455***
	(0.289)	(0.277)	(0.290)	(0.277)	(0.290)	(0.278)
$Branch_{i,t}$	-0.003***	-0.004***	-0.002***	-0.004***	-0.002***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$Deposit_{i,t}$	21.788***	26.816***	21.757***	26.789***	21.968***	27.001***
	(5.174)	(5.499)	(5.212)	(5.530)	(5.182)	(5.505)
$HHI_{i,t}$	-4.199***	-5.318***	-4.159***	-5.278***	-4.207***	-5.326***
	(1.143)	(1.212)	(1.140)	(1.200)	(1.138)	(1.205)
$White_{i,t}$	2.604***	1.527	2.783***	1.700*	2.635***	1.552
	(0.989)	(0.977)	(0.974)	(0.957)	(0.982)	(0.971)
$Female_{i,t}$	-0.655	3.480	-0.844	3.283	-0.676	3.465
	(8.853)	(9.312)	(8.910)	(9.387)	(8.854)	(9.323)
$Education_{i,t}$	7.115**	7.990***	7.518**	8.388***	7.427**	8.290***
	(2.984)	(2.331)	(3.027)	(2.367)	(2.966)	(2.311)
$Income_{i,t}$	0.267***	0.275***	0.264***	0.272***	0.263***	0.272***
	(0.022)	(0.020)	(0.022)	(0.020)	(0.021)	(0.019)
$Unemployment_{i,t}$	16.016**	15.985**	15.763**	15.737**	15.787**	15.752**
	(7.351)	(6.875)	(7.381)	(6.893)	(7.392)	(6.917)
$Senior_{i,t}$	-8.520**	-12.354**	-8.481**	-12.320**	-8.555**	-12.392**
	(4.185)	(4.857)	(4.290)	(4.980)	(4.178)	(4.853)
$Manufacture_{i,t}$	-0.009	-0.039	-0.008	-0.038	-0.009	-0.039
	(0.049)	(0.052)	(0.049)	(0.052)	(0.049)	(0.052)
$Information_{i,t}$	0.555	0.644	0.562	0.651*	0.562	0.650*
	(0.379)	(0.394)	(0.380)	(0.393)	(0.378)	(0.392)
$Flu_{i,t}^{ILI}$	-212.700	-169.219			-225.013*	-178.839
	(135.007)	(149.924)			(132.194)	(146.981)
$Flu_{i,t}^{POS}$			0.008	0.004	0.009	0.005
			(0.008)	(0.009)	(0.007)	(0.009)
Observations	70,991	70,991	70,991	70,991	70,991	70,991
Pseudo R^2	0.211	0.215	0.211	0.215	0.211	0.215
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 12: Social Connectedness, Loan Rates, and Loan Grades

This table presents the regression results for the following equation:

$$InterestRate_{i,k} = \beta RSCI_i + \gamma \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i indexes county and k indexes loan. The unit of observation is an approved loan. $InterestRate_{i,k}$ is the loan rate and is used as the dependent variable in Columns (1) and (2). In Column (3) and (4), the dependent variable is replaced with $LoanGrade_{i,k}$, which is a loan grade indicator variable with value of 35 for Grade A1 and decrease its value by 1 for each of the 34 sub-grades below A1. $RSCI_i$ is the social connectedness index for county i. $\boldsymbol{X}_k$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. Home ownership can be own, mortgage, or rent. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

	$InterestRate_{i,k}$		$LoanGrade_{i,k}$	
	(1)	(2)	(3)	(4)
$RSCI_i$	-3.613***	-0.792***	5.020***	1.180***
	(1.018)	(0.245)	(1.392)	(0.322)
$ln(Amount)_k$		0.629***		-0.814***
		(0.016)		(0.021)
$FICO_k$		-0.060***		0.083***
		0.000		0.000
$ln(Income)_k$		-1.201***		1.591***
		(0.022)		(0.029)
DTI_k		0.068***		-0.092***
		(0.001)		(0.002)
$ShortTerm_k$		-4.081***		5.688***
		(0.013)		(0.020)
$EmploymentLength_k$		-0.008***		0.011***
		(0.001)		(0.002)
$Population_{i,t}$		-3.690		8.296
		(3.854)		(5.377)
$Branch_{i,t}$		-0.000***		0.001***
		0.000		0.000
$Deposit_{i,t}$		-1.141***		1.533***
		(0.404)		(0.572)
$HHI_{i,t}$		-0.053		0.086
		(0.115)		(0.164)
$White_{i,t}$		-0.660***		0.908***
		(0.079)		(0.107)
$Female_{i,t}$		-1.254		1.726
		(1.662)		(2.234)
$Education_{i,t}$		-0.854*		1.109*
		(0.458)		(0.615)
$Income_{i,t}$		0.003		-0.004
		(0.003)		(0.004)
$Unemployment_{i,t}$		-1.080***		1.413***
		(0.352)		(0.492)
$Senior_{i,t}$		-0.185*		0.265**
		(0.095)		(0.128)
$Manufacture_{i,t}$		0.002		-0.003
		(0.003)		(0.004)
$Information_{i,t}$		-0.020		0.032
		(0.025)		(0.032)
Observations	1,535,725	1,535,201	1,535,725	1,535,201
Adjusted R^2	0.031	0.398	0.019	0.411
State FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Home Ownership FE	Yes	Yes	Yes	Yes
Loan Purpose FE	Yes	Yes	Yes	Yes

Table 13: Loan Performance and Social Connectedness

This table reports the coefficient estimates from the following logit regression:

$$BadPerformance_{i,k} = \alpha + \beta_1 RSCI_i + \gamma_1 Int_{i,k} + \gamma_2 \boldsymbol{X}_{i,k} + s_k + q_k + \epsilon_{i,k}$$

where i indexes county and k indexes loan. $BadPerformance_{i,k}$ equals 1 if at least one of the following statuses occurs on loan k within 24 months after loan origination: *Late (16-30 days)*, *Late (31-120 days)*, *Default*, *Charged Off*, and 0 otherwise. Panel A estimates the regression using the whole sample and Panel B estimates the regression for three separate sub-samples sorted by FICO score. $RSCI_i$ is the social connectedness index for county i. $\boldsymbol{X}_k$ is a vector containing a range of control variables, include loan characteristics, borrower characteristics, local demographic, economic, and social characteristic variables. s_i indexes state fixed effects and q_k indexes application month fixed effects. Standard errors are clustered at the state level and shown in parentheses. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***, respectively.

Panel A	$BadPerformance_{i,k}$		
	(1)	(2)	(3)
$RSCI_i$	-1.150***		-0.804***
	(0.308)		(0.202)
$InterestRate_k$		0.110***	0.110***
		(0.001)	(0.001)
$ln(Amount)_k$		0.192***	0.192***
		(0.006)	(0.006)
$FICO_k$		-0.006***	-0.006***
		0.000	0.000
$ln(Income)_k$		-0.126***	-0.127***
		(0.014)	(0.013)
DTI_k		0.009***	0.009***
		(0.001)	(0.001)
$ShortTerm_k$		0.163***	0.162***
		(0.007)	(0.007)
$EmploymentLength_k$		-0.018***	-0.018***
		(0.002)	(0.002)
$Population_{i,t}$		9.215**	4.426
		(4.230)	(3.503)
$Branch_{i,t}$		0.000	0.000
		0.000	0.000
$Deposit_{i,t}$		-0.970***	-0.799***
		(0.291)	(0.235)
$HHI_{i,t}$		0.081	0.077
		(0.054)	(0.049)
$White_{i,t}$		-0.071	-0.178***
		(0.055)	(0.059)
$Female_{i,t}$		-0.341	-0.673
		(0.605)	(0.759)
$Education_{i,t}$		0.110	0.270
		(0.174)	(0.182)
$Income_{i,t}$		-0.001	-0.001
		(0.002)	(0.002)
$Unemployment_{i,t}$		1.162*	1.065
		(0.692)	(0.662)
$Senior_{i,t}$		0.007	-0.104
		(0.138)	(0.138)
$Manufacture_{i,t}$		0.009*	0.010**
		(0.005)	(0.005)
$Information_{i,t}$		-0.102***	-0.119***
		(0.039)	(0.041)
Observations	1,535,723	1,535,199	1,535,199
Pseudo R^2	0.011	0.069	0.069
State FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Home Ownership FE	Yes	Yes	Yes
Loan Purpose FE	Yes	Yes	Yes

Panel B	$BadPerformance_{i,k}$		
	(1) Low FICOs	(2) Medium FICOs	(3) High FICOs
$RSCI_i$	-0.649***	-0.816***	-1.020***
	(0.202)	(0.223)	(0.351)
$InterestRate_k$	0.100***	0.109***	0.133***
	(0.001)	(0.001)	(0.002)
$ln(Amount)_k$	0.234***	0.161***	0.162***
	(0.009)	(0.008)	(0.010)
$FICO_k$	-0.007***	-0.007***	-0.003***
	(0.001)	0.000	0.000
$ln(Income)_k$	-0.085***	-0.116***	-0.180***
	(0.012)	(0.015)	(0.022)
DTI_k	0.012***	0.009***	0.005***
	(0.001)	(0.001)	(0.001)
$ShortTerm_k$	0.205***	0.167***	0.085***
	(0.010)	(0.010)	(0.016)
$EmploymentLength_k$	-0.017***	-0.017***	-0.021***
	(0.002)	(0.002)	(0.002)
$Population_{i,t}$	3.322	1.782	12.123*
	(3.311)	(4.101)	(7.070)
$Branch_{i,t}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$Deposit_{i,t}$	-0.641*	-0.815**	-1.139***
	(0.346)	(0.336)	(0.326)
$HHI_{i,t}$	0.115	0.030	0.101
	(0.073)	(0.079)	(0.066)
$White_{i,t}$	-0.211***	-0.166**	-0.116
	(0.060)	(0.072)	(0.111)
$Female_{i,t}$	-1.236	0.173	-0.964
	(0.881)	(0.792)	(0.875)
$Education_{i,t}$	0.218	0.363	0.195
	(0.181)	(0.275)	(0.242)
$Income_{i,t}$	-0.001	-0.001	0.001
	(0.002)	(0.002)	(0.002)
$Unemployment_{i,t}$	0.670	1.000	2.096***
	(0.591)	(0.774)	(0.791)
$Senior_{i,t}$	-0.105	-0.068	-0.155
	(0.157)	(0.184)	(0.185)
$Manufacture_{i,t}$	0.012**	0.009	0.008
	(0.006)	(0.006)	(0.006)
$Information_{i,t}$	-0.112**	-0.130***	-0.105**
	(0.049)	(0.039)	(0.047)
Observations	517,724	571,703	445,772
Pseudo R^2	0.049	0.053	0.074
State FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Home Ownership FE	Yes	Yes	Yes
Loan Purpose FE	Yes	Yes	Yes