

The Role of Family Networks in First-Credit Access

MIGUEL ANGEL CARPIO, LORENA KELLER, ALESSANDRO TOMARCHIO ^a

April 2023

Using a dataset of more than 38 million consumer-bank relationships in 518 districts of Southern Peru, we find that first-time borrowers receive credit and better loan terms from the bank where their families are more central. Our results are explained by informal-oriented banks using family ties as a strategy toward the unbanked. They give first credits to retain the relatives of the recipients as clients, and they also screen first-time borrowers using the credit behavior of their relatives.

^aCarpio: Universidad de Piura and Lima School of Economics, Lima, Peru. Email: miguel.carpio@udep.edu.pe. Keller: The Wharton School, University of Pennsylvania. E-mail: lorenak@wharton.upenn.edu. Tomarchio: University of Michigan and Central Bank of Peru. Email: alextom@umich.edu. The views expressed in this article are those of the authors and do not necessarily reflect those of the institutions of affiliation. We thank Jacelly Cespedes, Sean Flynn, Santiago Medina, Farzad Saidi, Pablo Slutzky, Dominik Supera, and Luke Taylor for very helpful comments. We also thank seminar participants at CEPR Webinar series in Finance and Development (WEFIDEV), European Winter Finance Meeting, Financial Management Association, Northeastern University Finance Conference, Swiss Society for Financial Market Research, Yale Junior Finance Conference, Naples Workshop on Networks and Development, Banco Central de Reserva del Perú, Superintendencia de Banca y Seguros, Caja Arequipa, Universidad de Piura, Wharton, and Asociación Peruana de Economía for great feedback. We are very grateful to Alexandra Heredia and Jorge Orozco for excellent research assistance. We are also very thankful to Vivien Cai for helping us out with the literature review. This project has greatly benefited from the financial support of The Jacobs Levy Equity Management Center for Quantitative Financial Research, Wharton's Dean Research Fund and Wharton's Global Initiatives Research Program.

1 Introduction

Expanding access to credit is crucial for improvements in welfare. The first loan an individual receives allows him to obtain credit reports that will be used to determine his borrowing terms with other financial institutions from then onwards. However, providing this first loan is risky for banks. They can reduce this risk, however, by making their lending decisions based on *hard information* (i.e., official documents on income, such as pay stubs and receipts of sales), which is captured in credit scores. In developing economies, however, this presents a problem. The majority of their population typically works in the informal sector and therefore do not have hard information to provide to banks. Even in developed economies, some segments of the population engage in informal activities and therefore lack, at least partially, hard information. The consequence is unequal access to credit.

Despite the importance of the first loan for the overall well-being of individuals, little is known about the strategies that banks use to reach those unbanked individuals engaging in informal activities. The ideal setting for learning about these strategies is the developing world, which has high levels of informal employment. In these countries, the supply of credit is typically segmented in two groups of banks, which we label as *formal-* and *informal-oriented*: formal-oriented banks are a small set of large, low-risk banks, that work in an environment with high barriers to entry, low competition and target formal workers; conversely, informal-oriented banks are a large group of small, risky banks, that operate in a highly competitive environment and target informal workers.

In this paper, we present empirical evidence that informal-oriented banks, as opposed to formal-oriented banks, actively use family ties as a strategy to approach first-time borrowers. We argue that, in a context of informal activities, these banks find value in the family link between a first-time borrower and the customers in their portfolio. We develop this argument in three empirical steps. First, we show that first-time borrowers are more likely to receive credit from the bank where their family is more influential. Second, we present indications that supply forces are important drivers of this effect, the most important of which is the fact that banks give larger loans and lower rates to first-time borrowers whose families are more influential. Finally, we focus on the possible sources of value that using family ties can bring to informal-oriented banks. The evidence

points to two novel mechanisms. First, banks provide credit to first-time borrowers that are family members of clients in order to retain the latter. Second, banks infer the ability to pay of first-time borrowers from their relatives’ reputation. We obtain these findings using data from Peru. As we will argue later, this country is a prominent example of the aforementioned segmentation, and it offers the ideal environment to identify family ties and to observe the strategies of banks towards the unbanked.

In the first step of our empirical approach, we show that first-time borrowers are more likely to receive credit from a bank where their family is more influential. As we explain later in more depth, our hypothesis is that banks, and more precisely informal-oriented ones, find value in their customers’ direct and indirect family connections. The more family connections the customer has, the more clients belonging to different families the bank can retain and the more prospective clients can screen.

Testing our hypothesis has three challenges. First, we need to map our hypothesis into an empirical framework. To do so, we develop a simple model that allows us to calculate the probability of a bank granting a loan to a first-time borrower as a function of family connectedness. Our derived expression maps this probability to a common indicator in network research to measure the influence of a family in a network. This indicator is *network centrality* (hereafter “centrality”). What this indicator does is to rank every family (a node in the network) based on its connections and on the connectedness of these connections.¹ The greater the value of this indicator, the more influential or more “central” a family is in the network.

The second challenge is to compute centrality in the data. This computation requires identifying the family networks of first-time borrowers, which are typically not known in the data. We overcome this challenge by using Peru’s credit registry, which includes family surnames. Following [Padgett and McLean \(2006\)](#), [Cruz, Labonne, and Querubin \(2017\)](#) and [Angelucci et al. \(2018\)](#), we define a *family* as the set of individuals that share the same surname within the district of residence. This allows us to build family networks. More precisely, we proceed in three steps. First, we use the credit registry to build a database for the universe of individuals in five regions of

¹This measure has been widely used in the literature of social connections. Examples are: [Alatas, Banerjee, Chandrasekhar, Hanna, and Olken \(2016\)](#); [Cruz, Labonne, and Querubin \(2017\)](#); [Banerjee, Chandrasekhar, Duflo, and Jackson \(2013\)](#).

Peru that accessed credit for the first time over the period 2010 to 2019 (approximately one million individuals). Second, since we want to study whether the importance of a family *within* a bank agency matters, we measure family influence at the level of each bank operating in the district of residence of the first-time borrower in the period prior to credit enactment. We call this set of banks the *consumption set*. Finally, we assign a centrality value to each individual by building family networks. These networks use the client portfolio of each district-bank combination. We conceive these portfolios as networks where each family is a *node*, and the joint occurrence of paternal and maternal surname in an individual's name is a *tie*. Based on these definitions, we graph full family networks and compute family centrality for each first-time borrower in each bank of his consumption set.

The third challenge is to estimate the causal effect. Our broad empirical strategy relies on comparing credit outcomes for the *same* individual across the alternative set of banks where his family is more or less central at the time he got his first credit. In other words, instead of comparing across individuals with different family connections, we compare the same individual on the date of his first loan across banks within the district of residence. By using individual fixed effects and exploiting family network variation within individuals across district–bank combinations, we mitigate the concern that our results are driven by differences across individuals. We also control for possible aggregate sources of variation coming from banks, districts and time by using a full set of interacted district, bank and quarter fixed effects.² Having said this, our results are robust without bank-district-quarter fixed effects. Similarly, we control for the most important characteristics of the families within each network, including the number of family members and the size and quality of the families' loan. However, our results do not depend on these controls.

Our results indicate that individuals have a higher probability of obtaining their first credit from banks where their family is more central. This is not only true for the discrete change resulting from having a relative in a bank's portfolio, but also for continuous changes in the level of centrality. The probability that an individual obtains his first loan from a bank increases by 0.3 percentage points if at least one member of his family is a client of that bank. This accounts for 12% of the average likelihood of obtaining a loan with a specific bank. More generally, a one standard-

²In effect, we have two possible sources of variation: One is within first-time borrowers, across banks; the second is within banks (within bank-district-time), across first-time borrowers.

deviation increase in centrality increases the likelihood of receiving a loan in that bank by 0.10 basis points, or 3% of the mean likelihood.

Having established that centrality matters for credit access, we proceed to the second step of our empirical approach. We present indications that supply forces are important drivers of our results. One could argue that our previous results might reflect individuals' preferences for choosing the same banks as their family members. For example, the first-time borrower might find banking in the same place as family members facilitates certain transactions. While this likely explains part of our results, demand forces alone cannot explain what we find. First, we find that all the centrality effect is driven by informal-oriented banks. If the centrality effect was fully explained by clients willing to carry out transactions with their relatives, we would find it across types of banks. Second, we find that our results are weaker for delinquent families.³ If our results were only demand driven, we would not expect there to be differences across levels of delinquency. Finally, we find centrality also affects the terms of the first loan. This points to a supply channel, since it is hard to believe that a borrower can influence the conditions of the credit he receives for the first time. To show that centrality affects the loan terms, we restrict our sample to those individuals who were extended credit by more than one bank within three months after receiving it for the first time. We use the same identification strategy (including individual fixed effects) and show that first-time borrowers receive better terms from banks where their families are more important, including loan size and interest rate. Taken together, these pieces of evidence suggest that our results are explained by banks' preferences, and not only by customers' preferences.

Finally, in the third step in our empirical analysis, we study the mechanisms underlying our results. We analyze why informal-oriented banks use family ties to make decisions when lending to first-time borrowers. Based on interviews with loan officers, we identify a group of potential mechanisms. First, because unlike formal oriented-banks, informal-oriented ones work in a very competitive environment, they may want to prevent their clients from migrating to other banks. They can do this by granting a first loan to a client's family member. We call this *perk* mechanism. Second, unlike formal-oriented banks, informal-oriented banks have almost no hard information

³We interact centrality with a dummy variable that indicates whether the share of relatives in delinquency is above 25%. We find this variable reduces the positive effect of family centrality by nearly half. Our results are also present with other levels of delinquency thresholds.

about the borrower. Hence, they can rely on non-traditional screening methods. If banks assume that the individual has similar characteristics than those of his family members, one of these screening methods could be as proxy for the credit quality of the individual with those of the family members to whom the bank is lending. We label this mechanism *reputation transfer*. Finally, other mechanisms already addressed in the literature could play a role. The client may help with screening by informing the bank about the personal information on his unbanked family member. He can also contribute to better monitoring and enforcement once this family member receives credit. All these identified mechanisms are examined empirically, one by one, to the extent that the data allows.

We expect the perk mechanism to be present if banks have seen that these practices, on average, increase customer retention. We analyze whether there is evidence for this. To do so, we expand our original database (at the first-time borrower—district—bank level) to analyze every direct relative within the district—bank portfolio. We use a comprehensive set of fixed effects, including individual fixed effects to control for potential confounders. We find that granting a loan to the first-time borrower is associated with direct relatives remaining longer as borrowers.

We do a similar analysis for the reputation transfer mechanism. We expect that banks will use this mechanism if they see that there is correlation between the payment behavior of direct relatives and the first-time borrower. Hence, we compute the correlation between the initial risk category of the first-time borrower's relatives in the loan portfolio, and the payment behavior of this subject after disbursement. We find a positive correlation, which suggests that the first-time borrower behaviour is related to that of his relatives.

We also analyze other mechanisms, but find no evidence in favor of them. For example, banks do not seem to ask their customers for private information about first-time borrowers, nor do they seem to be using family ties for better monitoring and enforcement. Having said this, the fact that we do not find evidence for these does not mean we rule them out.

Peru offers a great setting to provide empirical evidence for our framework. First, as 80% of its labor force works in the informal sector (Loayza, 2016), the above-mentioned supply segmentation is particularly marked. This makes the banks' strategies and their heterogeneities more visible. Second, because of their historical origins, the Peruvian surnames contain information on the

origin of the person (particularly within districts). Hence, one can identify family members through surnames. Moreover, in Peru, as in other countries that follow Hispanic naming conventions, individuals carry two surnames: the first is the father’s first surname and the second is the mother’s first surname. This allows us to build, for each bank, family networks of their client portfolios. Third, we have a detailed broad panel of consumer loans available, which allows us (i) to distinguish between a first loan in the banking system from other loans, (ii) to link individuals to families, and (iii) to get personal-level data from the whole financial system rather than a single bank.

To the best of our knowledge, our paper is the first to study how the banking industry approaches first-time borrowers, and the role that family networks play in this task. Our paper contributes to the literature linking social networks and credit access in developing countries. Previous research has three characteristics. First, it treats access to credit in general terms, without distinguishing between first-time and existing borrowers. However, it is valuable to do an exclusive analysis for first-time borrowers, because the first interaction between a borrower and a lender provides all banks with information about borrower’s creditworthiness through the credit reporting agencies. An exception is [Fisman, Paravisini, and Vig \(2017\)](#), who explore the client set of a bank in India. They show that cultural proximity between lenders and borrowers increases credit and reduces default. They analyze differences across borrowers that receive credit from the analyzed bank for the first time and the rest of borrowers, and do not find a difference between these subgroups. One possible explanation is that an individual who receives credit for the first time from a particular bank has most likely received credit from other banks in the past (i.e., this individual is not necessarily a first-time borrower, just a first-time customer at that lending institution), and therefore his credit history is available. We hope to contribute to the literature by focusing on first-time borrowers in terms of the whole credit market.

The second characteristic is that previous work uses data from one particular bank ([Banerjee et al., 2013](#); [Fisman et al., 2017](#)), or from surveys that do not identify the supplier ([Fafchamps, 2000](#); [Okten and Osili, 2004](#)). Although the literature provides valuable lessons, it does not show how banks differ in their approach to the unbanked. Our data at the individual $\times$ available bank help us to characterize the supply.

The third characteristic is that the literature on credit access focuses on social or political networks, for instance, whether borrowers and lenders attend the same sporting events, community gatherings, and religious celebrations (Fafchamps, 2000); whether they share ethnicity (Fisman et al., 2017); or whether lenders participate in community meetings (Okten and Osili, 2004). We focus on family, the most fundamental of societal institutions and the main source of connections in developing countries. An advantage of studying family networks relative to other social networks is that family is exogenous to an individual. It is given at birth rather than chosen as most other social networks are.⁴ We use family surnames to graph the full family network at the district $\times$ bank level, which allows us to compute the network position of all families using network centrality measures. These techniques are closest to Cruz et al. (2017), who use family networks to understand electoral outcomes.

The paper is divided into seven sections. Section 2 presents a simple conceptual framework of the interaction between family networks and credit access. Section 3 describes the context and the data sources for credit conditions, access to the first loan, and family networks. It also provides details on how we construct family networks and how the banking system in Peru is composed. Section 4 describes the empirical strategy for estimating the effect of family networks on first-credit access presents our estimations for these effects. Section 5 provides evidence that supply-side channels play an important role in explaining these effects. Finally, Section 6 presents suggestive evidence for the mechanisms described in Section 2. Section 7 presents several robustness checks. Finally, Section 8 concludes.

2 Framework

In this section, we argue that banks that target people working in the informal sector can find it beneficial to lend to first-time borrowers who have family ties with current clients of the bank. After developing our rationale for this argument, we use a simple model to explain how our hypothesis can be tested. We present tools used in network theory and explain how these allow us

⁴Okten and Osili (2004) also measures the number of economically active siblings and family wealth as a proxy of extension and quality of the family network.

to test whether family networks at the bank-level are linked to credit access and the terms of the first loan.

Our framework starts with the premise that there are two types of banks. These are the formal and informal-oriented banks. The former are large, low-risk banks that typically offer various products to their clients (such as portfolio investments) and do not focus exclusively on bank lending. The latter are the opposite: they are small, high-risk and focused on bank lending.

The difference in size and riskiness between these two groups of banks has various implications for their business models. First, while formal-oriented banks operate in a setting with significant barriers to entry, and hence with low competition, informal-oriented banks face fierce competition. Therefore, it is likely that informal-oriented banks work harder to retain clients through extra benefits and good service. Second, formal-oriented banks can borrow at significantly lower rates than informal-oriented ones and can therefore provide significantly lower rates to their customers. This in turn implies that banks can specialize in different types of clients who require different screening methods. While formal-oriented banks can lend to low-risk individuals who demand low rates, informal-oriented banks cannot. This makes informal-oriented banks target riskier clients, such as first-time borrowers working in the informal sector. Naturally, this means that unlike formal-oriented banks, which use credit score methods based on the observable characteristics of the loan applicant to estimate the probability of default a priori, informal-oriented banks need to use non-traditional screening methods to assess their clients' risk.

In this setting, which is compatible with [Broecker \(1990\)](#),⁵ family ties between a client and a first-borrower are significantly more valuable to informal-oriented than formal-oriented banks. There are many examples of the mechanisms through which informal-oriented banks may benefit from using family ties to approach first-time borrowers. First, informal-oriented banks can encourage customer loyalty and prevent clients from migrating to other banks by granting, as a perk, a first loan to a client's family members. This is important for these banks given the competitive environment in which they operate. We call this the *perk* mechanism. Second, without

⁵[Broecker \(1990\)](#) models the consequences of competition between banks on the intensity with which they screen borrowers. He finds that the rejection decision of a bank offering lower interest rates creates an externality on a bank offering higher interest rates: the former attract the best borrowers, leaving the low-quality borrowers to the latter. Our market segmentation adds to this model that the accepting decision of an informal-oriented bank creates an additional externality: first-time borrower start building up a credit history, which can be observed by competing lenders.

hard information at the individual level, informal-oriented banks could proxy the credit-worthiness of a prospective client by assuming that it resembles that of his family. Then, the bank can search within its current client network for how well their family members repay loans. We refer to this as the *reputation transfer* mechanism. In addition to these two mechanisms, the banking and social networks literature hints at some other. The client can help with screening by informing the bank about the personal information he has on his unbanked family member. He can also contribute to better monitoring and enforcement of the recently banked family member.⁶ He might even find it altruistic or convenient to assume part of the payment in case the family-tied client faces difficulties in repaying his loan.

While the examples above describe how lending to family members of current customers can benefit informal-oriented banks, it is likely that not all family ties are the same. The probability that a bank retains a current customer increases when a first-time borrower has closer family connections with him. We measure how close a first-time borrower is to a current client by the family distance in a network. For example, consider the client portfolio of bank *A* shown in Panel A of [Figure 1](#). In this figure, each node is a family who is a client of the bank and the links represent ties between families.⁷ We argue that if bank *A* lends to an unbanked individual of family 1, the bank will have higher chances of retaining current customers which are only one step away from family 1 than those further away.⁸ This same logic applies to the other benefits besides those of client loyalty (e.g., inferring the credit-worthiness, improving screening, monitoring and enforcement of the prospective client).

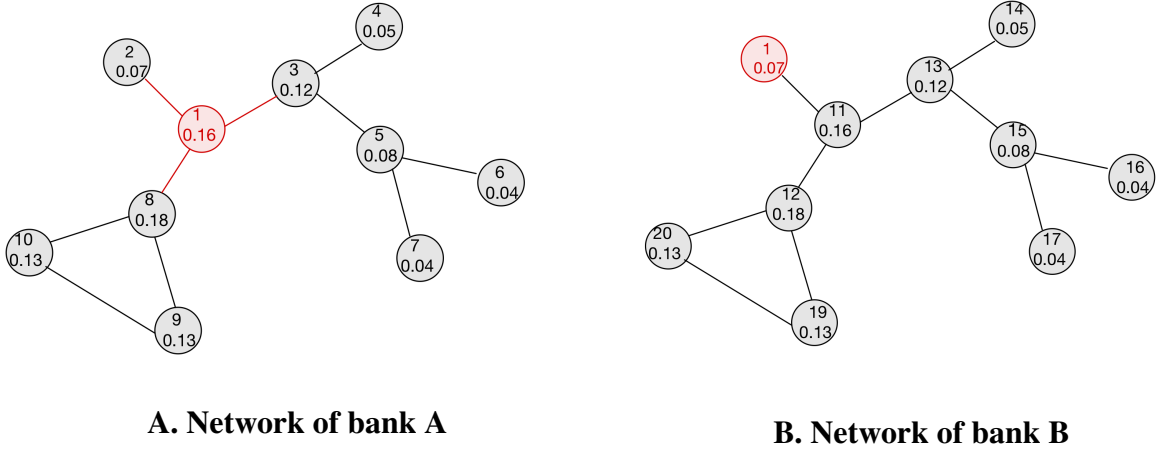
To help map our hypothesis relating family networks to credit access to our empirical analysis, we formalize our claims using a simple model. In this model, bank *b* is evaluating whether to lend to a set of prospective clients. To do so, the bank scores each prospective client *i* based on two dimensions. The first score, σ_{bi} , ranks the individual based on the personal characteristics that are unrelated with family ties. The second score, $score_i^b$, is determined by the sum of the

⁶In a context of group lending in Peru, [Karlán \(2007\)](#) finds that individuals with stronger social connections to their fellow group members have higher repayment due to successful monitoring.

⁷[Section 3](#) will explain in detail how we define a tie and how we construct family networks.

⁸That is, the bank has higher chances of retaining families 2, 3 and 8 (1 step away) are higher than retaining families 4, 5, 9 and 10 (2 steps away). At the same time, the bank has greater probability of retaining families 4, 5, 9, 10 than families 6 and 7 (3 steps away).

Figure 1: Network illustrations



Note: This figure shows two examples of a bank's network. In both panels, each node is a family and it is identified with a number, which is on the top of each node. The center of the node contains the eigenvector centrality of each node.

benefits that the bank expects to get from direct and indirect ties between i and every client j of its portfolio.⁹ The longer the distance l between a prospective client i and current client j , the lower the probability that the bank derives a benefit from that connection. Using network theory notation, we can express the value that bank b assigns to the family relationships of individual i with its current client base (i.e. $score_b^i$) as:

$$score_b^i = \sum_j \sum_{l=1}^{\infty} a_{ij,l} \times \alpha_{ij}^l \times benefit_b \quad (1)$$

where α_{ij}^l denotes the probability that bank b derives a benefit of $benefit_b$ from the family relationship of length l between prospective client i and current client j ; and $a_{ij,l}$ are the elements of the matrix A^l that records the number of connections of length l between families i and j .¹⁰ Note

⁹For example, in Figure 1, Panel B, when Bank B lends to an unbanked member of family 1, the bank benefits from both, the direct connection between family 1 and 11, and from the indirect connection that family 1 has with other families in the network. An example of an indirect connection is the relationship with family 12. This comes from having a relationship with family 11, which has a relationship with family 12.

¹⁰In network theory, a network is represented by the *adjacency matrix*, A . This matrix contains the links between different nodes in a network. The element a_{ij} equals one if nodes i and j are connected and zero otherwise. As we will explain in more detail in Section 3, this matrix notation will allow us to represent mathematically the family networks in each bank in our data. The elements of A^l , denoted $a_{ij,l}$ capture connections of length l between families i and j . Considering all possible connections between the unbanked individual belonging to family i and family j , the value

that the first summation fixes a particular family j and checks all the paths between the family of the unbanked individual i and j , while the second summation sums across all possible families j in the bank network. Whether there is a relationship between i and j and the length of this relationship will be considered in $a_{ij,l}$.

The expression of Equation (1) corresponds to a measure used in network theory to compute the importance of a node, called *Katz centrality*. It determines the relative importance of a node based on two categories. The first is the number of links it has to other nodes. The second is the relative importance of those connections. That is, it recognizes that not all links are equally important. While it weighs distant connections less than closer connections, being connected to a node which itself has many connections is more important than being connected to a node with no other connections. For example, the middle of each node in Figure 1 records the centrality value of the node. Notice that in Panel A, family 1 and family 5 both have 3 direct connections, but because family 1's connections are better connected than those of family 5, family 1 has a greater centrality score.

The way centrality is computed neatly captures the properties of $score_i^b$. First, $score_i^b$ should be higher for prospective clients i with more family connections. The greater the number of connections, the more clients belonging to different families that the bank can retain, screen and monitor, hence, the more that the bank benefits from lending to i . Second, $score_i^b$ should also consider the connectedness of family ties. The better connected are those family ties, the greater the benefits for the bank.¹¹

Having described the two scores σ_{bi} and $score_b^i$, we now describe how the bank combines these two scores to guide its lending decisions. The bank approves applicants with scores above a cutoff

of the family ties between i and j is given by $\sum_{l=1}^{\infty} a_{ij,l} \alpha^l$. For example, consider that the network of bank B is only given by the subset of nodes 1, 11 and 12. In that case, the value of the tie between 1 and 11 in the notation we are showing here would be $1 \times \alpha^1 + 0 \times \alpha^2 + 0 \times \alpha^3 + \dots$. The first term comes from family 1 and family 11 having a relationship of length 1. Hence the matrix element $a_{1,11} = 1$. However, we also sum over the connections of length 2, 3 until infinity that there could be between 1 and 11. All of those equal zero because the length between family 1 and 11 is 1. Repeating this exercise for all possible families j in the network with which the unbanked individual i (fixed) can have a relationship, gives the expression in Equation (1).

¹¹To visualize this, consider family A has one connection, family B, but family B 10 connections. If family B is retained because the bank lends to a member of A, the bank has higher probability of retaining family B's 10 connections as they can stay in the bank because B stays. Now, if B did not have any further connection, the value for the bank of lending to a member of A decreases as it loses its chances of indirectly retaining B's 10 connections. This is why in Panel A, Figure 1, family 1 has a greater centrality than family 5.

c and rejects applicants with scores below. More precisely, bank b grants a loan to i if and only if $score_i^b + \sigma_{bi} > c$. Letting Φ be the cumulative distribution function of σ_{bi} and assuming it is symmetric, it follows that the probability that bank b grants a loan to i is given by:

$$\begin{aligned} P(b \text{ grants a loan to } i) &= P(\sigma_{bi} > c - score_i^b) \\ &= 1 - \Phi(c - score_i^b) \\ &= \Phi(score_i^b - c) \end{aligned} \tag{2}$$

Hence, by denoting the Katz centrality for individual i in the client portfolio of bank b as K_{bi} and simply combining Equations (1) and (2), we obtain that the probability that the informal-oriented bank b grants a loan to the unbanked prospective client i is a function of K_{bi} .¹²

$$\begin{aligned} P(b \text{ grants a loan to } i) &= \Phi(K_{bi} - c) \\ &= f(K_{bi}, c) \end{aligned} \tag{3}$$

Equation (3) summarizes the mapping between our framework and our empirical analysis. Since banks find family ties between a prospective client and current clients valuable, the probability of an individual i taking a first loan in a bank is a positive function of the centrality of his family within the network of the client portfolio of bank b . The function can be made linear by simply assuming that Φ is a uniform distribution. In our empirical analysis, the main dependent variable is the probability that an unbanked individual has access to a loan, and the main independent variable is centrality. In our empirical analysis, we use eigenvector centrality, because it is the most commonly used in the literature (e.g., Banerjee, Chandrasekhar, Duflo, and Jackson (2013); Cruz, Labonne, and Querubin (2017)), and it has the same properties as Katz centrality. The only difference is the weight set on distant versus close connections.

To be more specific, our empirical analysis computes the eigenvector centrality measure to each family node and compares credit outcomes for the same individual. Figure 1 illustrates this

¹²Our model is similar to Cruz, Labonne, and Querubin (2017), which is applied to a political economy context. In their model, mayoral candidates with higher eigenvector centrality receive more votes because they can be approached by a larger number of voters through the use of fewer intermediaries.

strategy. Every panel represents the family network within the client portfolio of two different banks, A and B . Consider again the unbanked member of family 1 who is about to receive credit for the first time. Notice that family 1 has more connection in bank A than in bank B . It is also better located in A than in B , that is, the families connected to family 1 in each network are more important in terms of connections. This means that the centrality of family 1 is greater in A (0.16) than in B (0.07). We predict that our unbanked member of family 1 is more likely to borrow from A than from B .

3 Context and data

In this section, we present the data we use to construct the key variables in our empirical analysis. [Section 3.1](#) discusses the data and construction of our main dependent variables, which are credit outcomes. [Section 3.2](#) discusses the importance of family in our area of study and describes the construction of our main independent variable, family centrality. Finally, [Section 3.3](#) provides the institutional background of the banking system in Peru.

3.1 Access to first credit

To obtain our credit outcomes, we merge two datasets: (i) the *electoral roll*, and (ii) the *credit registry*. Our identifier is the unique individual identification number. The electoral roll provides us with the district of residence of the universe of individuals who are 18 years old and older. We use this to build family networks at the bank-district level. The credit registry provides us with the credit outcomes. Next, we describe these sources in more detail.

Electoral Roll The electoral roll is the list of potential voters for the 2011 elections for President and Vice Presidents, which contains the identification numbers, names, surnames, and districts of residence of the voting population. The population eligible to vote is the universe of Peruvian citizens who are 18 years of age or older, regardless of whether they exercise their right to vote. The National Office of Electoral Processes provided this confidential database for five out of the 25

regions of Peru: Ayacucho, Apurimac, Arequipa, Cusco and Puno. These regions cover a fifth of the national territory and comprises the Southern Andes, as shown in [Figure A.1 Online Appendix](#). The number of people in this database is 3.05 million.¹³

Credit Registry The credit registry is a confidential database of the Peruvian Banks Supervisor (SBS). It includes detailed information on the universe of loan balances for all individuals and all banks in Peru at the monthly level.¹⁴ It includes the size of the loan balance, currency (Peruvian *soles* or U.S. dollars), indicators of loan quality (e.g., delinquency rate), and the monthly accrued interest charged on the loan. Importantly, it also includes the identification number of the loan holder, which allows us to merge this database with the previous one. In order to reduce the computational load of working with monthly balances, we migrated from monthly to quarterly data.

Merged data To build our final dataset, we proceed in three steps. In the first step, we create a sub-sample that contains only the individuals we are interested in studying, which are first-time borrowers between from December 2010 to December 2019. Out of the 3.05 million on the electoral roll, we end up with 952 thousand individuals. For these first-time borrowers, we only keep their credit information for the quarter in which they received their first loan. In other words, an individual only appears on the date that he first received credit, so there is no time-series within individuals.¹⁵

In the second step, we go back to the merged database to determine the consumption set of each first-time borrower. In our baseline regressions, we define this consumption set as the collection of banks with at least one customer having a loan balance greater than zero, in the same district of residence, in the period prior to the first-credit access. We consider several alternative definitions

¹³The political organization of the country includes (in descending order in terms of size) regions, provinces and districts. We have data for 52 out of the 196 provinces (27%), and 518 out of the 1,869 districts (28%). Panel B of [Figure A.1](#) shows how districts are divided in our sample. Districts that are region capitals or provincial capitals have much better access to public and private goods than other districts.

¹⁴The SBS periodically receives information from banks, consolidates it, and makes it available to banks, private risk information centers, and the Central Bank. Lenders can also access the information through individual inquiries.

¹⁵For the majority of cases, this effectively means that the individual only appears once in our dataset, paired with the bank who gave the individual its first loan. The exception occurs when the individual obtains his first loan with many banks at the same time.

for the consumption set in [Section 7.2](#). For instance, we define it as either (i) banks of the same type (i.e., formal or informal-oriented) based on the ex-post decision of the individual, or (ii) banks where each individual has family members. In all cases, the evaluation is carried out in the period prior to the first-credit access.

Finally, in the third step, we use the set of available banks in each district, to expand the subsample built in the first step. Now, our unit of analysis is the combination individual–available bank in the district of residence. Again, there is no time-series within individual, but the individual repeats itself n times in the same period, where n is the number of bank alternatives. This data has 38.4 million observations.

With our final dataset, we build our dependent variables. First, when studying the impact of family networks on credit access, our dependent variable is a dummy that equals one if the individual obtained his first credit from that bank. Otherwise, it is zero. Since our unit of analysis is individual–available bank in the district of residence, we can have a within individual identification strategy, where we can compare the probability of getting a loan for the *same* individual (at the same time) *across* banks. Second, to evaluate credit conditions, our main dependent variables are the logarithm of the loan and the interest rate. As we will explain later, given these variables are only defined conditional on the individual getting a loan, our *within* individual analysis constrains our sample to those individuals who got their first loan from two institutions simultaneously.

3.2 Family centrality

Throughout history, family ties have been strong in the Peruvian Andes. Individuals have strong ties with both, immediate and extended family members. In the Andes, these extended families have historically formed communities called *ayllus* ([Murra, 1975](#); [Wachtel, 2017](#)), whose members not only consume together, but also help each other in terms of production and exchange of goods and services ([Alberti and Mayer, 1974](#), p. 14). Part of the persistence of these communities is explained by the lack of solid institutions and effective social security systems in these areas. Hence, the family is arguably the strongest social group and the main source of social capital.

We have defined family as the set of individuals that share the same surname within the district of residence. The origins of surnames in Peru can be found in the Spanish conquest in the sixteenth century (Carpio and Guerrero, 2021). In the course of converting indigenous people to Christianity, the Spanish practiced mass baptisms, which conferred first names and surnames to the baptised. The first name was typically Hispanic, surname typically ethnic. The assigned surnames contain information on the origin of the indigenous person, *ayllu* or “caste” (Garcilazo de la Vega, 1609, p. 28). As very common family names are not as prevalent in the Southern Andes as in modern regions of Peru, and there has not been much migration to these areas, surnames are strongly correlated with family membership. The literature has also shown surnames to be similarly informative in Philippines (Cruz, Labonne, and Querubin, 2017) and areas of Italy (Barone and Mocetti, 2021; Gagliarducci and Manacorda, 2020). However, to mitigate concerns regarding the introduction of noise coming from very common surnames, and hence, mistakenly linking people as family, we also show that our results are robust to eliminating the most common surnames.

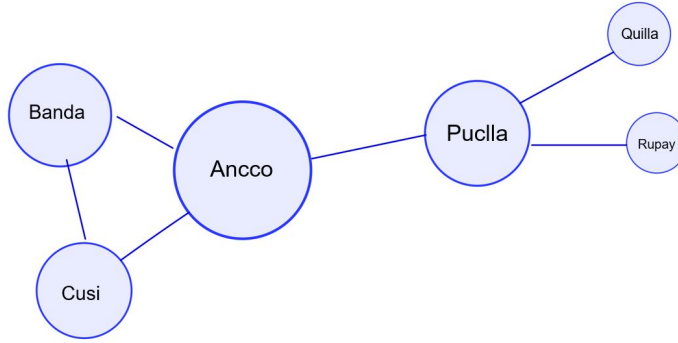
Surname conventions in Peru allow us to link individuals with their paternal and maternal families. This happens because an individual is assigned two surnames: One paternal (the father’s first surname), and one maternal (the mother’s first surname). Take, for example, an individual named Diego Ancco Puclla. Diego is the first name; it is discretionary and given by the family. Ancco and Puclla, Diego’s surnames, are family names and are predetermined. Ancco is the father’s (first) surname. Puclla is the mother’s (first) surname. The joint occurrence of paternal and maternal surname is thus strongly correlated with a tie between two families and a tie between two historical castes. This is especially the case within districts.

We exploit these naming conventions to generate family networks at the bank-district-quarter level and construct our main independent variable. This variable is the eigenvector centrality, which measures the number of related families who also use the bank, but also weights each of those families by their importance in terms of connections and network position.

We construct eigenvector centrality for each individual obtaining his first loan in each bank-district pair. To do so, we proceed in three steps. In the first step, we exploit naming conventions to measure large scale family networks in each bank-district-quarter. For each bank, in each district

and each quarter, we consider all of its current borrowers (i.e., the bank's entire lending portfolio). Using all of these borrowers, we graph the family network structure. An example of this structure for bank b in district d in quarter q is given in Figure 2. This structure could arise when the bank is lending to our fictitious Diego Ancco Puclla. For this reason, we see a tie between Ancco and Puclla in the figure. The bank lends to Diego's cousins, Felipe Puclla Quilla and Paola Puclla Rupay on his mother's side, and to Diego's cousins Ricardo Ancco Banda and Mario Cusi Ancco on his father's side. Moreover, Mario Cusi Ancco has a cousin whose surname is Cusi Banda who also borrows from the bank. As we can see, the nodes, which are families, are tied by marriage linkages.

Figure 2: Example of family network



Note: Example of a family network, in a district d for a bank b , at the time t . Each node (surname) represents a family. The size of each node is proportional to its eigenvector centrality. Ancco's centrality is greater than Puclla's. The centrality of Banda and Cusi is the same. Quilla and Rupay also have the same centrality.

In the second step to building the centrality for each individual obtaining its first loan, we compute the centrality of each surname (node) in every bank, district and quarter. In our example for bank b , district d and quarter q , we estimate that the centrality of node *Ancco* is equal to 0.58, while the value of node *Puclla* is 0.42. Although Ancco and Puclla both have two direct connections, Ancco has a higher centrality than Puclla because its connections are themselves more connected. We replicate this computation for every bank-district in each quarter of our sample.

Finally, the last step to obtain the centrality of a new borrower who is not currently part of the bank's portfolio is to use the centrality of the surnames computed for each bank network and assign the centrality values to each individual obtaining its first loan in each bank-district pair. We do this based on the unbanked individual's surname. However, since every individual has two

surnames (the father’s first surname and the mother’s first surname), we follow [Cruz, Labonne, and Querubin \(2017\)](#) and assign the maximum centrality value to each individual. The assumption is that if individuals have a high degree of centrality in one network but not in the other, they may get the possible benefits of being central as long as they are central within one of the networks. Our results are robust to changing the assignment of centrality and considering, for example, the average centrality of an individual.

In addition to centrality, we compute a few additional variables to characterize the network. In particular, the number of relatives in the network, the relatives’ credit size and the number of relatives in the network with good credit quality. We use these variables to control for family attributes such as size and wealth.

3.3 Institutions

We have 78 banks in our sample. A first group of banks is oriented towards the formal sector. It is made up of 18 banks, all of them authorized to carry out a large number of financial activities, and are active on a national scale. A second group, oriented towards the informal sector, is made up of 60 banks, authorized to carry out a limited number of financial activities and are usually regional in origin.¹⁶

We compare these two types of banks. [Table 1](#) presents descriptive statistics by type in every column. Panel A shows information from our merged database, which is composed of the universe of individuals with a positive balance in the different quarters. Each row presents the quarterly average of a specific variable. First, the formal-oriented are larger than informal-oriented banks. The average number of clients is 24,922 for the formal-oriented banks and 16,196 for the informal-oriented banks. Loans are also larger: The average credit size is 15,677 *soles* for the former and 8,729 *soles* for the latter (about 4,020 and 2,240 U.S. dollars, respectively). Second, the formal-

¹⁶The regulations of the SBS classify institutions into Banking Company, Financial Company, Municipal Savings and Loan, Municipal Savings Association, Entity for Small and Micro Business, Rural Savings and Loan, and Specialized Company. Our first group is made up exclusively of Banking Companies, while the second one is composed of other types of institutions. However, we consider three Banking Companies within the second group as their mission explicitly mentions the objective of reaching unbanked individuals: Banco Agropecuario, Banco Azteca and Mibanco.

oriented banks do not incorporate new clients into the credit market as the informal-oriented do: in a quarter, on average, the former extends credit to only 242 first-time borrowers, while the latter extends it to 462 individuals. This is consistent with their substantially different approaches to attracting unbanked clients. Third, perhaps contrary to what one might expect, the average percentage of individuals with good credit quality is similar in formal- and informal-oriented banks (76% and 81%, respectively).

Table 1: Descriptive Statistics by type of institution: Quarterly Averages

	Formal-oriented	Informal-oriented	Difference (1)-(2)	
	(1)	(2)	Coef.	Tstat
Panel A: Popultaion				
Number of banks	18	60		
Credit amount (Million S/)	390.6902	141.3841	249.3061**	2.2649
Number of clients	24,921.9826	16,195.8895	8,726.0930	1.0583
Mean loan	15,676.5308	8,729.6289	6,946.9020	1.2055
Clients <i>bancarized</i> : number	242.1536	462.4617	-220.3081	-0.9726
.....: %	0.6098	2.1521	-1.5423***	-3.5462
Clients with good credit quality (%)	76.1014	81.3889	-5.2875	-0.8820
IHH Index	2,140.1744	1,051.4673	1,088.7071	1.5869
Panel B: ENAHO sample				
Number of banks	13	51		
% main occupation is formal	69.1259	28.9073	40.2186***	6.2754
Avg. HH annual net income per capita	16,154.6150	9,197.9547	6,956.6603***	6.6854
% with high school or more	84.4553	58.8854	25.5699***	3.5784

Note: This table presents descriptive statistics of institution's characteristics by type. Each row presents the quarterly average of a specific variable. Panel A considers as banks' clients the universe of individuals with a positive balance in different quarters. Panel B considers as banks' clients the subsample of credit holders surveyed by the ENAHO between 2010 and 2019. Column (1) considers institutions that are formal-oriented while column (2) institutions that are informal-oriented.

In Panel B, we merge our credit data with *The National Household Survey* (ENAHO), a yearly large survey that tracks living conditions and poverty, to compare the target segment by type of bank. We were able to merge the ENAHO with the credit data because we had access to a confidential version that includes the individual identifier.¹⁷ Customers differ markedly by type of bank. For instance, consistent with our classification, the average percentage of clients whose main occupation is formal is 70% for formal-oriented banks and only 29% informal-oriented banks. Along the same lines, the former cater more to people with higher household incomes (16 versus 9 thousand *soles*) and completed high school (85% versus 59%).

¹⁷The year of the interview is not considered for the merger. In a sense, all waves of the ENAHO are treated as a single cross-section.

We finish by giving some context regarding fintechs. In our period of analysis, fintechs in Peru had not evolved to the same degree as in other countries in the region. According to [Marín, Barragán, and Zaballos \(2014\)](#), Peru exhibits one of the lowest fixed broadband penetration rates per household in Latin America, well below the regional average (18% versus 30% in 2012). There is also a gap between cities (particularly Lima) and the interior districts.¹⁸ However, penetration of mobile services is higher, reaching 80% for Peru and 58% in rural areas.

4 Credit access

In this section, we test one of the predictions of [Section 2](#), which is that an individual has higher probability of getting his first loan in a bank where his family is more central. We test the remaining predictions in the following sections.

Estimating the effect of family networks on credit access is challenging. Banks' decisions to lend to first-time borrowers depends on an individual's characteristics that can be correlated with those of his family. Hence, comparing credit access across individuals who belong to different families could mistakenly attribute the results to family networks when instead, these were a result of differences across individuals.

Our baseline regression specification addresses this challenge by exploiting the fact that individuals have different centralities across institutions in the same district at the time of their first loan. We compare the effect of family centrality *within* individual and *across* banks, rather than *across* individuals within a bank. Notice we do not attempt to compare whether an individual whose family is more important in a particular bank has a greater probability of accessing a first-loan than an individual whose family is less important for that bank. Instead, we take the same individual and ask whether he is more likely to receive his first-credit from a bank where his family is more important rather than from a bank where his family is less important. This allows us to account for all fixed individual characteristics that might have confounded the effect of eigenvector centrality, such as wealth and financial literacy. More precisely, we estimate:

¹⁸Municipalities with access to fixed broadband greater than 10% represent 2% of the total, and municipalities with access to broadband less than 1% are 85%.

$$y_{dbit} = \alpha \text{ centrality}_{dbi,t-1} + \beta X_{dbi,t-1} + \gamma_{d \times b \times t} + \eta_i + \mu_{dbit} \quad (4)$$

where y_{dbit} is a dummy indicating whether the individual i , resident of district d and recipient of the first loan in period t , receives credit from bank b . Our key treatment variable is $\text{centrality}_{dbi,t-1}$, the eigenvector centrality of family i in the customers' portfolio of the bank b in district d in the period prior to the individual's decision. $X_{dbi,t-1}$ are *district* $\times$ *bank* $\times$ *family*-specific controls one period before the individual belonging to that family gets his first loan. $\gamma_{d \times b \times t}$ is a complete set of interactions of district, bank and time fixed effects. η_i is the set of individual fixed effects. Finally, μ_{dbit} is the usual idiosyncratic error term and standard errors account for potential correlation within banks-districts.¹⁹ For ease of comparison, we multiply the dependent variable by 100, so the coefficient of interest is interpreted directly as basis points.

In this regression, our within-individual analysis is the result of using individual fixed effects, η_i . The time dimension captures that different individuals receive their first loan at different points in time. There is no time-series dimension within-individual as its time is fixed to the date t in which the individual got his first loan. Hence, in this case, we only need individual fixed effects and not individual-time fixed effects.²⁰

Moreover, because one could be concerned that our results are affected by: (i) economic shocks that vary over time; (ii) particular strategies of certain banks; (iii) particular strategies of bank agencies in certain districts; and (iii) differences across districts, we also include in the regression the interaction between bank-district-time-fixed effects. This should absorb any potential confounder that changes across time, banks, districts, or a combination of these. As an example, these control for the case where banks, or even bank agencies in a district, have many clients and therefore individuals have large centrality in that bank. Because of the banks' size,

¹⁹Clustering at the bank level also yields similar results.

²⁰Similarly, because there is only one individual per district, we do not have individual variation to control at the district level (i.e. we do not need individual-district fixed effects). With this in mind, before setting individual fixed effects, there were only two sources of variation at the individual level. These were: (i) those common across banks and (ii) those that vary by bank, such as individual preferences to choose a particular bank, or banks' preferences for certain individuals. We control for the first using individual fixed effects but we still have variability at the individual-bank level. Our results suggest that the variation at the individual-bank level seems driven by bank choices for particular individuals (i.e. the most central) rather than preferences of individuals for certain banks.

these banks could also be lending to many new borrowers. We control for this scenario because all individuals in that bank-district will face a similar condition. Therefore, the fixed effect at the bank-district level absorbs this source of variation. With our fixed effects, the only source of variation we are allowing is one that varies by individual-bank but is unrelated to a shock in a particular time or a general policy of the bank, or bank office in a district. An example for a bank credit supply that varies at the individual level is banks' preferences to lend to specific borrowers, such as the most central.

The exhaustive set of fixed effects removes the most important sources of endogeneity. The remaining concerns would be an omitted variable that changes at the same level as the main explanatory variable (i.e., district–bank–time–surname level), which is correlated with both centrality and credit access. An example one can think of is a number of banks located in a specific district that decide to focus on wealthier families of specific surnames. We try to mitigate this possibility by controlling for family-specific characteristics, such as family wealth, size and credit quality. These are the set of controls in the set $X_{dbi,t-1}$. The proxies we use for wealth, size and credit quality are the family's total amount of credit, the number of family members in the network (as in [Cruz, Labonne, and Querubin \(2017\)](#)) and family members with good credit, respectively. All of these controls are measured the quarter before the individual takes out his first loan.

Notice that an advantage of choosing family rather than any other social network is that family is exogenous to an individual. An individual in Peru cannot change their surname, except in rare cases. It is fixed at birth. The family network within an institution is also exogenous to the individual. Because we compute the centrality of an individual's family the quarter before he gets his first loan, his affiliation with a particular bank does not affect his centrality at the time of choosing banks.

Our benchmark results of the effect of family networks on credit access are shown in [Table 2](#). The first three columns show the estimates of Equation (4) for different definitions of centrality. In the first column, our main explanatory variable is a dummy that takes a value of one if our measure of eigenvector centrality is positive. This occurs when the individual has family in the bank. The probability that an individual obtains his first loan from a bank increases by 0.30 percentage points if at least one member of his family is client of that bank. Given that the probability that an

individual has his first loan with a specific bank is 2.5%, our estimate is economically large, as it accounts for 12% of the average likelihood of obtaining a loan with a specific bank.

Table 2: The effect of centrality on credit access

	Credit given			
	Main results			Informal-oriented interaction
	(1)	(2)	(3)	(4)
Centrality > 0, [t-1]	0.2999*** (0.0101)			
Centrality < Indiv Median Centrality, [t-1]		0.1487*** (0.0104)		
Centrality > Indiv Median Centrality, [t-1]		0.4884*** (0.0173)		
Centrality [t-1]			0.0993*** (0.0050)	0.1307*** (0.0062)
Centrality [t-1] × B: Is formal-oriented				-0.1322*** (0.0108)
Indiv FE	Yes	Yes	Yes	Yes
Dist × Bank × Time FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Y Mean	2.53	2.53	2.53	2.53
N indiv	952,136	952,136	952,136	952,136
N	38,378,450	38,378,450	38,378,450	38,378,450

Note: This table shows the effect of family centrality on credit access for unbanked individuals. The dependent variable is a dummy indicating whether the individuals i , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. The explanatory variable differs by columns. Column (1) uses a dummy equal to 1 when centrality is positive and 0 otherwise. Column (2) uses two indicator variables that use the median centrality for each individual conditional on this being positive. With this median, the variables are dummies indicating whether the centrality of a bank for an individual is above or below the median. Column (3) uses the continuous measure of eigenvector centrality. To simplify interpretation, the eigenvector centrality has been rescaled to have mean 0 and standard deviation 1. Column (4) has as explanatory variables (i) the eigenvector centrality, and (ii) an interaction between the eigenvector centrality and a dummy on whether the bank b is formal-oriented (=1) or informal-oriented (=0). Standard errors are in parenthesis. They are clustered at district × bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second column shows that the probability of obtaining a first loan with a bank increases monotonically with centrality. This suggests that we are effectively capturing the effect of family networks on credit access. Conditional on centrality being positive, we compute the median centrality for each individual. With this median, we split the positive centrality cases of Column (1) into two dummies: one where centrality is below the median, and one where it is above the median. Relative to the case where the individual has no family in the bank, having family members with only weak family ties (below the median), increases the probability of getting a loan in that bank by only 0.15 percentage points. This coefficient is more than three times as large in banks where the individual's family ties are the strongest (above the median). In these banks, family networks account for 20% of the mean likelihood of obtaining the first loan in that bank.

More generally, the third column shows the continuous measure of centrality. To simplify interpretation, we follow [Cruz, Labonne, and Querubin \(2017\)](#) and we normalize centrality to have a mean of zero and standard deviation of one. A one standard-deviation increase in centrality increases the likelihood of receiving a loan in that bank by 0.10 basis points, or 3% of the mean likelihood.

In the Online Appendix, [Table A.I](#) introduces the fixed effects and controls of Equation (4) gradually. Most of the fixed effects decrease the magnitude of the coefficients without changing their significance.²¹ The inclusion of the set of family-specific controls yields essentially the same results.²²

Our results are robust to a wide set of alternative specifications. These are presented in detail in [Section 7](#). Among other tests, our specifications are robust to (i) using alternative centrality definitions, (ii) dropping common surnames, (iii) excluding married women who may have changed their last name, and (iv) using a different set of bank alternatives to build the relevant “consumption set”.²³

5 Supply Channel

The last section established that family networks matter in access to credit. These results could be driven by credit supply channels (e.g., bank’s strategies toward recruiting unbanked individuals) or credit demand channels (e.g., individuals wishing to transact with their relatives). In this section, we find that credit supply channels are important drivers of our results. Note that this does not

²¹The largest decrease in the coefficient of interest is caused by the introduction of fixed effects at the district-bank level, going from 0.60 in Column (5) to 0.17 in Column (6). This change is consistent with our results being driven by only one type of bank: the informal-oriented.

²²The coefficient is 0.10 in Columns (9) and (10), which suggests that family-specific characteristics might not be playing a role in our results.

²³One might suggest the use of an explanatory variable that values not only family ties, but also the particular importance of each family, such as wealth or credit balance. However, our research objective is to observe the effect of the family ties, precisely without other variables interfering. In any case, we have estimated our benchmark regressions by weighting each observation by the credit balance that the family of the first-time borrower holds in each district-bank. The results, available upon request, are very similar.

mean that we rule out demand channels. Demand channels are likely also present but we have no evidence to provide in favor of them.

To show supply channels are important, we proceed in three steps. First, we show our results are driven by informal-oriented banks. Credit demand channels would suggest that our results should be present across all bank types. Second, we show that our results are significantly weaker for families with a bad reputation. Finally, and most importantly, we show that family networks matter for the terms of the loans, such as interest rates and credit amounts. Demand forces are also unlikely to be behind this effect, as first-time borrowers cannot change the conditions of the loan they receive. Hence, while demand forces can be explaining part of our results, our results indicate that supply channels play a very important role.

5.1 Asymmetry between formal and informal-oriented banks

A starting point of our framework was that formal-oriented banks do not find as much value into family ties among their clients as do informal-oriented banks. For instance, as formal-oriented banks offer lower interest rates, they do not need to encourage clients' loyalty by granting loans to their relatives. Similarly, they might not use family ties to infer the credit-worthiness of a prospective client because they use credit score methods. Consequently, we expect the effect of family networks on credit access not to be driven by formal-oriented banks.

Column (4) of [Table 2](#) displays the results of adding to the baseline regression (Column (3)) an interaction between family centrality and a dummy that takes the value of one if the bank is informal-oriented. As expected, we observe that all of the results of family centrality are driven by informal-oriented banks. The net effect of family centrality on access to credit in formal-oriented banks is nil. On the contrary, a one-standard-deviation increase in family centrality in informal-oriented banks increases the likelihood of obtaining a loan by 5% of its mean likelihood.

The results on the heterogeneity between formal- and informal-oriented banks is evidence in favor of the claim that our results cannot be driven only by credit demand channels. If the results were driven only by credit demand forces (e.g., individuals wishing to transact with their relatives), we would expect that these hold for all types of banks.

5.2 Centrality and bad family reputation

Based on the same considerations affecting the heterogeneities between formal- and informal-oriented banks, we expect that if our results were only driven by demand channels, the delinquency rate of the family should not play a role. In both cases, an individual preferences for transacting with family members would drive them to borrow from the same bank as their family, regardless of their family's delinquency rate. However, if the supply side plays a role, then the effects of centrality on credit access will be weaker for highly delinquent families. The rationale lies in two mechanisms discussed in [Section 2](#). These are the perk and reputation mechanisms. From the perk side, one would expect that informal-oriented banks are less interested in retaining families who are delinquent. From the reputation transfer side, if these banks assume first-time borrowers can have similar delinquency rates than family members, one would expect that these banks will be less inclined to lend to first-time borrowers whose families are delinquent.

We test whether our results on credit access differ across differences in family delinquency. The results are shown in [Table 3](#). Our measure of bad reputation of a family in a bank is the share of family members (clients of the bank) who are late with loan payments.²⁴ We then interact our centrality measures with different variants of this variable. Column (1) repeats the effect of a dummy that takes a value of one if our measure of eigenvector centrality is positive, which we have shown in [Table 2](#). Column (2) adds a dummy that also takes a value of one if our measure of bad reputation is above the median of the distribution.²⁵ Having at least one family member in the bank's portfolio decreases from 0.32 to 0.25, that is, decreases by a fifth, the probability of an individual obtaining his first loan from a bank. Column (3) adds a dummy that takes a value of one if our measure of bad reputation is above the arbitrary cutoff of 25%.²⁶ We use this variable as an

²⁴In the Peruvian regulatory framework, loans are classified into five categories according to their level of risk. The "normal" category indicates that a loan is up to date with its payments. Our variable of family reputation is the percentage of family members whose category is not "normal".

²⁵The share of family members with delayed loan payments can only be computed for the alternatives where the first-time borrower has family members (i.e., *centrality* > 0). We assign zero to our measure of bad reputation if the individual does not have family members in the bank's portfolio. Conceptually, we are equating not having relatives in the bank with having relatives in good standing. [Section 7](#) shows the result are robust to restrict the sample to individuals with family members.

²⁶11% of the observations is above this cutoff. This percentage rises to 24% if observations without family members in the bank are removed.

indicator of which are the worst families. The centrality effect decreases from 0.35 to 0.17, that is, it decreases by a half.

Table 3: The effect of centrality on credit access, by family's reputation

	Credit Given						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Centrality > 0, [t-1]	0.2999*** (0.0101)	0.3213*** (0.0107)	0.3491*** (0.0112)				
Centrality > 0, [t-1] × D-B-I: Above median rel w/ bad reputation[t-1]		-0.0715*** (0.0125)					
Centrality > 0, [t-1] × D-B-I: Above 25% rel w/ bad reputation[t-1]			-0.1798*** (0.0146)				
Centrality [t-1]				0.0993*** (0.0050)	0.1125*** (0.0057)	0.1121*** (0.0056)	0.1092*** (0.0058)
Centrality [t-1] × D-B-I: Ratio rel w/ bad reputation					-0.0007*** (0.0001)		
Centrality [t-1] × D-B-I: Above median rel w/ bad reputation[t-1]						-0.0308*** (0.0075)	
Centrality [t-1] × D-B-I: Above 25% rel w/ bad reputation[t-1]							-0.0425*** (0.0084)
Indiv FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist × Bank × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	2.53	2.53	2.53	2.53	2.53	2.53	2.53
N indiv	952,136	952,136	952,136	952,136	952,136	952,136	952,136
N	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450

Note: This table shows the effect of family centrality on credit access by family's reputation. The sample is composed of unbanked individuals. The dependent variable is a dummy indicating whether the individual i resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. Each column presents a different regression. Column (1) uses only a dummy equal to 1 when centrality is positive. Column (2) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (3) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Column (4) uses only continuous centrality. Column (5) uses the ratio of relatives with bad reputation. If the individual has 0 relatives the ratio is 0. Column (6) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (7) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Standard errors in parenthesis. They are clustered at district×bank level. We control for the number of relatives in the network in $t-1$, the relatives' credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We conduct the same analysis with our continuous measure of centrality. Column (4) simply repeats the benchmark result of Table 2. Column (5) adds an interaction of centrality with our measure of bad reputation. The coefficient is negative and significant, which implies that the centrality effect decreases with bad reputation. To help reference the size of this decrease, Column (6) adds instead an interaction of centrality with the dummy equal to one if our measure of bad reputation is above its median, and Column (7) adds an interaction of centrality with the dummy equal to one if it is above 25%. In both cases, the centrality effect decreases from roughly 0.11 to 0.07. Note that since bad reputation does not completely offset the effect of centrality, there are other forces that still allow banks to derive utility from family ties.

Section 7 presents robustness checks for this table, including an alternative definition of family reputation (instead of the share of family members with delay in credit payments, the share of their

loan portfolio), and a recalculation of the table eliminating the observations of individuals with zero relatives in the bank. The results are similar in both cases.

There are three takeaways from [Table 3](#). First, it suggests supply side channels can be important. Second, it is consistent with banks deriving less benefits from family ties when these families have bad reputation. Third, it suggests that perk and reputation transfer mechanisms could be some channels present in explaining our results. We provide further tests of these channels in [Section 6](#).

5.3 The effect of centrality on credit conditions

Next we explore whether family networks improve credit conditions as we would expect if supply side channels dominate. To do so, we employ a similar empirical strategy to the one used to test the effects of family networks on credit access. We still use the main regression specification described by Equation (4). That is, we use individual fixed effects, $\text{district} \times \text{bank} \times \text{time}$ fixed effects and family controls. The only difference is that now the dependent variable is either the logarithm of the credit amount, the interest rate²⁷ or the quality of the loan.

There is a caveat, however, when analyzing credit conditions. As in the case of access to credit, applying individual effects is crucial to preventing our results from being explained by differences across individuals. However, unlike the case of access to credit, where the dependent variable contains information for each individual in each bank available, credit conditions are only available in the banks where the individual accepted a loan. Therefore, by using individual fixed effects we are limiting our sample to those individuals who obtained their “first loan” in more than one bank simultaneously.

There are three problems with using individual fixed effects and hence a sample of only those individuals who got their first loan in more than one bank simultaneously. The first is that this clearly shrinks our sample compared to the *unrestricted* sample we had when computing our results for credit access. To limit the reduction in our sample, we consider all individuals who

²⁷The interest rate on the loan is not readily available. To obtain this variable, we follow [Céspedes \(2017\)](#) to compute an implied interest rate. The implied rate results from dividing the accrued interest by the loan balance. We work with annualized rates and, as in [Céspedes \(2017\)](#), we find that our implied rate closely matches the aggregate rates published by the SBS.

received their first loans with more than one bank within a window of three months. This *restricted* subsample includes 115,337 individuals and 250,011 observations (individuals-institutions). Our assumption is that a three-month window is not large enough to generate credit information capable of changing the credit conditions offered by the second and subsequent banks.^{28 29}

The second problem is that our sample differs from the population studied in [Section 4](#). More precisely, we find that those getting more than one loan at the same time are more likely to live in urban areas, to work in the formal sector, to have higher levels of education and earn higher income (see [Table A.II](#) in the Online Appendix). However, even if the characteristics of the individuals in this subsample are not the same as those of those who received only a single first loan, it can still be informative to determine how family centrality affects credit conditions in an environment in which we use individual fixed effects and our results cannot be confounded with other individual characteristics.

The last problem is that we find that nearly 80% of individuals taking two or more loans at the same time take these from the same type of bank. That is, either all from informal-oriented banks (69% of the individuals) or all from formal-oriented banks (9% of individuals). Hence, there is very little variability across institution types that could be used to study whether our results differ between formal or informal-oriented. This was not the case when studying credit access. For credit access, as the individual's consumption set was determined using all formal and informal-oriented banks (although we do robustness checks over this consumption set) in which people in their district had a loan, we had variability within individuals across bank types. This allowed us to add an interaction term with a dummy of formal-oriented banks to determine whether our results were driven by formal or informal-oriented institutions. To address this issue and still get a sense of whether our results on credit conditions might be driven by informal-oriented banks, we use sample splits in our analysis of formal versus informal oriented banks. We study how family networks affect credit conditions in areas where informal-oriented banks dominate, in contrast to areas where formal-oriented banks dominate.

²⁸At the end of each month, the SBS receives from the banks information on the loan balances corresponding to the end of the previous month. For instance, a loan received in January will appear in the registry at the end of February, and the first payment condition (i.e., whether the client met the first monthly payment) will appear at the end of March.

²⁹In [Section 7.3](#) we show that our results are robust to controlling for the order in which the first-time borrower received the loan within the 3-month window.

Table 4 presents the results on credit conditions. Panel A presents the results of the complete sample of districts. We find that individuals obtain better credit conditions in banks where their families are more important (i.e., more central). A one-standard-deviation increase in family centrality is associated with a 1.4% increase in credit and 1.5 percentage points decrease in interest rates.

Table 4: The effect of centrality on credit conditions

	Log (credit) (1)	Interest rate (PEN) (2)	Credit quality (6m later) (3)
Panel A: All Districts			
Centrality [t-1]	0.0135 (0.0083)	-1.5299** (0.7040)	0.0066 (0.0042)
Y mean	7.1568	48.5109	0.7971
N indiv.	82,472	57,757	57,367
N obs.	178,635	123,532	123,724
Panel B: Districts with larger % of formal-oriented banks			
Centrality [t-1]	0.0035 (0.0124)	-0.4430 (1.0049)	0.0067 (0.0074)
Y mean	7.1744	47.8327	0.8010
N indiv.	31,098	21,794	20,134
N obs.	66,312	45,892	42,860
Panel C: Districts with smaller % of formal-oriented banks			
Centrality [t-1]	0.0221* (0.0122)	-2.8953*** (1.0112)	0.0064 (0.0049)
Y mean	7.1432	48.9580	0.7943
N indiv.	50,512	35,234	36,546
N obs.	110,363	76,022	79,313
Indiv FE	Yes	Yes	Yes
Dist $\times$ Bank $\times$ Time FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: This table shows the effect of family centrality on credits conditions. All regressions have individual fixed effects, district $\times$ bank $\times$ time fixed effects and controls. The sample differs by panel. The first panel considers individuals that got their first loans in more than one bank within a three month window and the banks in which they got their first loans. The second and third panel use subsamples. The second panel includes only the individuals that live in districts with larger share of formal-oriented institutions available than the median district. The third panel includes those living in districts with smaller share of formal-oriented institutions available than the median district. The explanatory variable is eigenvector centrality (rescaled to have mean 0 and standard deviation 1). The dependent variable differs by column. Column (1) uses the logarithm of the loan amount. Column (2) uses the interest rate of loans made in the Peruvian currency (PEN). Column (3) uses a dummy equal to 1 if the loan has a good credit quality six months later. Standard errors are in parenthesis. They are clustered at district $\times$ bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel B and C present results of the subsamples that result from splitting the sample by districts. Panel B shows the effect of family centrality on credit conditions in districts where the share of formal-oriented banks is above the median. Panel C shows the complement. Individuals living in districts dominated by formal-oriented institutions receive only slightly better conditions on their first loan in banks where their family is more central. The size of the loan is practically the same as in banks where the individual is less central. Although the rates in the bank where the individual

is one standard deviation more central are on average 0.4 percentage points lower, this result is statistically insignificant. In contrast to these results, individuals living in districts where the share of informal-oriented banks is above the median (Panel C), receive 2.9 percentage points lower rates in the bank where they are more central. They also receive loans 2.2% larger in those banks.³⁰

6 Possible mechanisms

Section 2 presents some mechanisms that informal-oriented banks may use in the framework of their strategy of using family ties when they approach first-time borrowers. In this section, we analyze some of them empirically, to the extent our data allows. Banks are likely to use some of these mechanisms simultaneously. We will not present a quantification of which is more important, although the evidence will be clearer with some than with others. Other mechanisms could also be present but we cannot test them empirically.

Perk We have argued that banks may find convenient to build customer loyalty by granting a loan to a first-time borrower who is a family member of a current client of the bank. This would alleviate the risk of losing them as clients. We expect banks to follow this strategy if they see that it works.

We create a new dataset to determine whether the perk mechanism is at play and allows banks to retain current clients related to first-time borrowers when banks grant the first loan to these clients' unbanked relatives. The starting point is our original database of first-time borrowers on the date they received their first loan. Recall that it includes the set of bank alternatives these borrowers had at that time (their consumption set).³¹ For each of these bank alternatives, we include the direct relatives the borrower has in the bank with positive credit (henceforth *family clients*), where

³⁰We have also explored the effect of centrality on loan terms by family's reputation, as we did in Table 3. In particular, we regress our dependent variables on our continuous centrality variable and an interaction of this variable with a dummy that takes one if our measure of bad reputation is above the arbitrary cutoff of 25%. When the dependent variable is credit size, we obtain a positive coefficient for centrality and a negative one for the interaction. When it is interest rate, the coefficient is negative for centrality and positive for the interaction. Although these are the expected signs, the coefficients are not statistically significant, which we attribute to the three problems explained above.

³¹Notice that, for each bank-district, we include first-time borrowers who received a loan from that bank and those who did not (although they received from other bank).

direct relatives include every individual with the same surname who has positive credit in the combination of bank and district. Hence, we go from a first-time borrower—district—bank database of 38 million observations, to a first-time borrower—district—bank—family client database of 677 million observations. We then add in this dataset, for each family client, an indicator of whether the family client continues to be a client of the bank (i.e., whether his credit balance continues to be positive).

We use this data to estimate the effect of granting a loan to a first-time borrower on the retention behavior of family clients. The coefficients of a naive estimation would capture the comparison between family clients whose relative received his first-time loan in the bank and those other family clients whose relative did not receive his first-time loan in that bank.

However, this estimation is subject to biases. First, we could be comparing across very different family clients and our results could be driven by channels outside perk. For example, it may be that banks grant loans to relatives of family clients whose business is growing and need loans. Then, the retention of bank clients could be driven by differences across the growth rate of family clients' businesses rather than the perk channel. Using family-client fixed effects mitigates this problem. It allows us to compare the same family client across banks, and therefore, measure whether the behavior of the same family client is associated to the bank in which his relative received his first loan.

A second bias in the naive estimation is that we would be comparing the behavior across family clients who are related to first-time borrowers with very different characteristics. These characteristics could both drive banks to lend to some of these first-time borrowers while at the same time spurring changes in the behaviors of family clients. This could be the case if banks lend to successful first-time borrowers who employ family members. Family clients working in this business could decide to continue borrowing to invest in the family business, while at the same time, encouraging banks to grant a first loan to the successful first-time borrower. Adding first-time borrower fixed effects alleviates this concern as this allows comparison between family clients within each first-time borrower.

Following the previous discussion, we use both first-time borrower and family client fixed effects. These fixed effects exploit the fact that: (i) the same individual may be a client of two

different banks comprising the consumption set of the same first-time borrower; and, (ii) the same individual may be related to more than one first-time borrower. Moreover, we also use $\text{bank} \times \text{district} \times \text{quarter}$ fixed effects to control for variation at the bank-district-quarter level. In sum, we estimate:

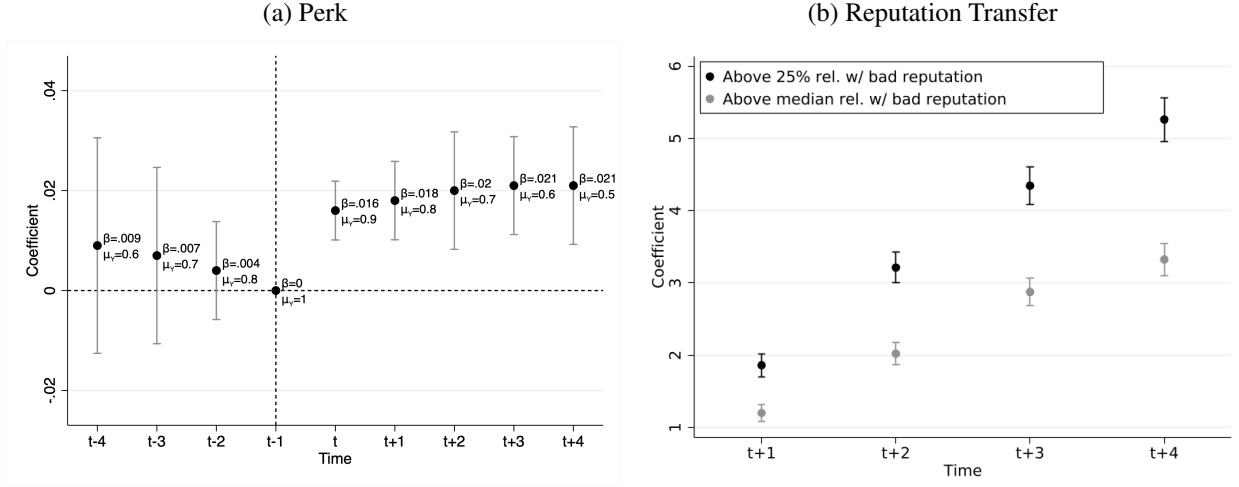
$$y_{dbij,t+x} = \alpha \text{credit-given}_{dbijt} + \gamma_{d \times b \times t} + \eta_i + \sigma_j + \mu_{dbijt} \quad (5)$$

where $y_{dbij,t+x}$ is a dummy equal to one if the loan balance in bank j of family client j , who is related to the first-time borrower i , resident of district d and recipient of the first loan in period t , is greater than zero. The dependent variable of the family client is evaluated a few periods after t , such that $x = 1, 2, 3, 4$. Our key treatment variable is $\text{credit} - \text{given}_{dbijt}$, which is equal to one if the bank grants credit to the first-time borrower i . As in previous regressions, $\gamma_{d \times b \times t}$ is a complete set of interactions of district, bank and time fixed effects.³² η_i is the set of first-time borrower fixed effects and σ_j is the set of family client fixed effects. Finally, μ_{dbijt} is the usual idiosyncratic error term and standard errors account for potential correlation within banks-districts.

Figure 3a plots $\hat{\alpha}$ estimated from Equation (5) various quarters before and after t , which is when the bank lends to the first-time borrower. We see that granting a loan to a first-time borrower raises the probability that his relatives keep a positive loan balance (relative to other banks in which the family client had loans). We see an increase of 1.8 percentage point in the probability one quarter after the bank grants the credit to the first-time borrower. This effect keeps increasing as time passes by, up to 2.1 percentage points four quarters after the bank lends to the first-time borrower. Since the average probability that the family client stays in the bank drops as time goes on (this is a natural consequence of the expiration of the loans), the effect becomes more important as times goes on. Figure 3a shows these results were not present before the bank lent to the first-time borrower. This alleviates any concern that our results could be spurious and driven by previous trends.

³²Recall that “time” does not imply a time series. Our “time” variable just controls for the date in which the first-time borrower got his first loan. We only have one observation per first-time borrower, which is the initial date of his loan.

Figure 3: Possible mechanisms



Note: Panel(a) shows the effect of chosen institution by a first-time borrower on the permanence of the family member in the institution. The dependent variable indicates whether the family member j resident of district d has an active credit in institution b . The explanatory variable *credit-given* is a dummy indicating whether the first-time borrower obtained his/her first loan at the same institution in which the relative had an active loan. Panel(b) shows the coefficient obtained by regressing a dummy, indicating whether the first-time borrower is delinquent in periods $t+1$, $t+2$, $t+3$ and $t+4$, on the reputation of the family in $t-1$. We use "Above 25% relatives w/ bad reputation" y "Above median relatives w/ bad reputation" as a proxy of family reputation in the bank.

These results suggest that the informal-oriented banks' strategy of granting first loans to the relatives of their clients could have the expected positive effects of retaining clients. This is consistent with banks finding family ties valuable. [Section 7](#) shows our results are robust to controlling for the specific relationship between a first-time borrower and the family client, by using first-time borrower $\times$ family client fixed effects. They are also robust to controlling for the relationship between family clients and banks by adding family client $\times$ bank fixed effects. We obtain even larger coefficients.

Reputation transfer We have argued that an informal-oriented bank may infer a first-time borrower's probability of repayment from the risk rating of his family members in the bank's portfolio. They would do this if they have seen that delinquency ratios are positively correlated among family members. Consistent with this hypothesis, [Table 3](#) shows that the effect of centrality on the probability that an individual receives credit from a bank decreases when his family has a bad reputation with the bank.

We now use our original database to show that there is a positive correlation between the risk category of the relatives of a first-time borrower in the loan portfolio one period before the

disbursement and the payment behavior of this subject after disbursement. Figure 3b shows the results.³³ The coefficient is 0.04, 0.07, 0.10 and 0.11 for the first, second, third and fourth quarters after disbursement, respectively. We focus only on these periods because the sample size drastically shrinks as we consider longer periods.³⁴ This evidence suggests that banks would do well to use family prestige to predict the payment behavior of the first-time borrower.

Screening device In Section 2, we also argued that banks might ask clients for personal information about their unbanked family members who are looking to borrow. Table 5 provides suggestive evidence that this channel does not explain our findings. We merge our credit registry with ENAHO to obtain the personal characteristics of the individuals who obtained their first credit. Because only a sample of individuals is surveyed, our sample for this regression includes approximately 13,000 individuals and 500,000 observations. The coefficients of the interactions of family centrality with personal characteristics such as age, sex, education and income level, do not yield any statistically significant results.

Our results suggest that banks that use family networks to expand their credit portfolio use these networks with all types of customers, regardless of personal characteristics. Hence, current clients would not be revealing personal information about the prospective client to the bank.

Monitoring and enforcement In Section 2, we argued that banks may find family ties between their clients and an unbanked individual valuable because clients can contribute to better monitoring and enforcement of the recently banked individual. This could be particularly important if family members do not want to their banking relationship to be hurt when other members of their family do not pay back their loans in the bank where they also clients. In this case, family members might monitor other members in the same bank more closely and help them out in paying their loans for the specific bank where they are also clients.

³³We regress the percentage of relatives with bad credit (i.e., the number of relatives with credit with at least one day of delay in payments over the number of relatives in the bank) on a dummy variable equal to one if the first-time borrower is one or more days late in paying his loan.

³⁴Borrowers disappear from the database when they fully pay off the loan, and also when they default on payments for several consecutive periods. In the latter case, under conditions established by the supervisor, the banks declare the loan as uncollectible and write off the amount from their financial statements.

Table 5: The effect of family centrality, by individual characteristics

	Credit Given						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Centrality[t-1]	0.0962*** (0.0250)	0.1097*** (0.0347)	0.1005*** (0.0377)	0.0907** (0.0376)	0.0930*** (0.0278)	0.0973*** (0.0355)	0.1321*** (0.0635)
Centrality[t-1] × Is above median age[t-1]		-0.0383 (0.0512)					-0.0443 (0.0546)
Centrality[t-1] × Is female			-0.0172 (0.0516)				-0.0153 (0.0525)
Centrality[t-1] × Complete high school or more				0.0033 (0.0504)			-0.0080 (0.0593)
Centrality[t-1] × Complete higher education					-0.0036 (0.0670)		-0.0017 (0.0734)
Centrality[t-1] × Is above median HH income p.c. capita						-0.0118 (0.0516)	-0.0134 (0.0542)
Indiv FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist × Bank × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	2.47	2.46	2.46	2.46	2.46	2.46	2.46
N indiv	13,687	13,210	13,238	13,122	13,122	13,238	13,118
N	558,617	539,378	540,516	535,497	535,497	540,516	535,327

Note: This table shows the effect of family centrality on credit access by individual characteristics. The sample is composed of unbanked individuals that were interviewed by the ENAHO between 2010 and 2019. The dependent variable is a dummy indicating whether the individuals i , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. Each column shows a different regression. Column (1) shows the baseline regression. The following six columns show regressions that have as explanatory variables (i) the eigenvector centrality (rescaled to have mean 0 and standard deviation 1), and (ii) an interaction between the eigenvector centrality and an individual characteristic. The characteristics differ by columns. Column (2) uses a dummy equal to 1 if the individual i is older in $t-1$ than the median individual, and 0 otherwise. Column (3) uses a dummy equal to 1 if the individual i is female; otherwise its value is 0. Column (4) and (5) use a dummy equal to 1 if the individual i has completed at least high school or higher education, respectively; otherwise it is equal to 0. Column (6) uses a dummy equal to 1 if the annual net income per capita of individual i 's household is larger than the median household, and 0 otherwise. Finally, column (7) shows a regression that includes all the interactions. Standard errors are in parenthesis. They are clustered at district×bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

However, Table 4 does not provide evidence in favor of this mechanism. If the family network actually contributes to better monitoring and enforcement, then centrality should have an effect on the quality on the credit. Column (3) uses a dummy equal to one if the loan has a good credit quality six months later, that is, if the shares of the credit have been paid punctually during the six months after its granting. Panel A presents the results of the complete sample of districts. We find a positive coefficient, but it is small and not statistically significant. The results are the same in the subsamples. Hence, although we have no evidence in favor of the proposed channel, we cannot rule it out.

7 Robustness

7.1 Centrality construction

The construction of the main explanatory variable has obviously been subject to arbitrary decisions. Throughout this subsection, we make alternative decisions and show that our results are robust. We also mitigate some concerns related to the use of surnames to build family networks.

Alternative centrality measures We have chosen the conventional eigenvector centrality. [Table A.III](#) in the Online Appendix shows that family networks matter for other types of centrality, such as degree centrality, closeness centrality, betweenness centrality.³⁵

Paternal and maternal surname Each individual has a paternal surname and a maternal surname and, therefore, can belong to two families or nodes in the network of a bank. We have followed [Cruz, Labonne, and Querubin \(2017\)](#) and used the maximum eigenvalue centrality of the two surnames. An alternative measure would be to take the average between the two last names. The second row of [Table A.IV](#) in the Online Appendix shows our results are robust to using the average eigenvector centrality between the two surnames instead of the maximum eigenvector centrality.

Surnames changes by marriage Married females in Peru may voluntarily add their husband's surname to their own. Before 2009, this worked by replacing their second surname with the first surname of their husbands preceded by the word “*de*”, which means “of”. From 2009 on, a third surname is added, preceded by “*de*”. As the 2011 electoral roll (from which we take the surnames) likely includes modified maternal surnames,³⁶ we replicate our main results but this time excluding individuals whose maternal surname is preceded by “*de*”. The third row of [Table A.IV](#) in the Online Appendix shows that our results are not altered.

Informational content of surnames The stochastic process behind surname dynamics means that a small number of the most frequent surnames make up a large percentage of the population,

³⁵*Degree centrality* measures how connected a node is, that is, the number of direct links. *Closeness centrality* measures how easily a node can reach other nodes. As its name suggests, it is a measure of how close it is from other nodes. *Betweenness centrality* measures the share of times that a node falls between two other nodes if one wants to get from one node to another.

³⁶The percentage of maternal surnames preceded by “*de*” is 11.28% in the case of women and 0.24% for men.

while a large number of low-frequency surnames are found among the population. A concern in this context is that very frequent surnames in a district do not actually show a family relationship between those who have them. Therefore, we exclude the most common surnames within districts. To do so, we compute the percentage of individuals with the same surname within every district. We rank the combinations surnames-districts according to this share. We then run regressions excluding individuals whose paternal or maternal surname is very common within their district. [Figure A.2](#) in the Online Appendix plots the coefficients after dropping the most common surnames in each district, gradually from just one to the ten most common. The effects are very stable.

7.2 Consumption set

The baseline construction of the consumption set assumes that if a bank is available for someone in the district, it is also available for someone else in the same district. We make various robustness checks with alternative definitions and we find that our results are robust.

Positive centrality One might worry that our results are an artifact of comparing access between a bank where an individual has no family members to one where it has some. In Column (4) of [Table A.IV](#) in the Online Appendix, we restrict the consumption set to banks where each individual has family members, and show this is not the case.

Market segmentation One could be concerned that, due to the strong market segmentation, the consumption set of an informal individual is composed exclusively of informal oriented banks. We then build an alternative consumption set. If the individual took a loan from the informal-oriented bank, we assume that the relevant choice of bank was only the informal-oriented one. We do the same for formal-oriented. The results are shown in the second row of [Table A.V](#) in the Online Appendix. We see that our results remain nearly unchanged.

Small banks One possible concern is that banks with few customers in a district may have low centrality levels for a first-time borrower, and also not be actual alternatives for them. Hence, the positive relationship between centrality and the probability of granting first credit would be artificial. This claim is false in the context of our specification. Our comprehensive set of fixed effects includes the district-bank combination, which controls whether alternatives are more or less

attractive to first-time borrowers as a whole. However, we deal further with this concern by trimming the consumption set. In particular, we exclude banks that do not exceed a specified portion of the market share, defined as the number of residents of a district with positive loan balance from a specific bank over total number of residents with positive loan balance. In [Table A.V](#) of the Online Appendix, Rows (3) to (6) shows the results for cutoffs of 0.5%, 1%, 2% and 3%. The coefficients monotonically rise with respect to the baseline, which alleviates the concern. Moreover, we take another approach by restricting the consumption set to banks with at least one office in the district of residence. Row (7) shows the coefficients also increase with this definition.

Overall, alternative definitions of the consumption set lead to larger centrality effects. We prefer the widest definition as it keeps the largest number of individuals in the analysis, and it makes the baseline results the most conservative ones.

7.3 Robustness checks for specific tables

Centrality effect by family’s reputation In [Table 3](#), we have shown that banks avoid relying on family networks to give first-credit to individuals for the worst families in their portfolios. Our measure of a family’s bad reputation in a bank is the share of family members (clients of the bank) who have been late with loan payments. However, one can argue that a better measure of family reputation is the size of the compromised loans. [Table A.VI](#) in the Online Appendix uses an alternative measure of bad reputation: the share of the loan portfolio of those family members with some level of delay in the loan payments. The results are basically the same, although the drop in the centrality effect is a little more dramatic in the event that the percentage of the compromised portfolio exceeds 25%, which seems quite intuitive.

In [Table 3](#), we have also assigned a value of zero to the measure of bad reputation if the individual does not have family members in the bank’s portfolio. Conceptually, we were equating not having relatives in the bank with having relatives in good standing. [Table A.VII](#) in the Online Appendix restricts the sample to individuals that have family members in the bank’s portfolio. The number of individuals drops from 952 to 919 thousand, while the number of observations drops from 38 to 18 millions. However, the results remain qualitatively the same.

Credit conditions One could worry that our findings of [Table 4](#), showing that first-time borrowers receive better terms in the bank where their family is more central, could be driven by the order in which loans were granted rather than centrality. This concern can arise because we combine loans within a quarter as if they had been given simultaneously, on the assumption that a quarter is not enough time to build credit history. However, in reality, one loan could have been granted a month or two before. If within that quarter, the first loan the individual received is where his family is most central, the first bank could have given better conditions to the borrower because at that point the borrower had no other debt. However, when the second bank, where his family happens to be less central, evaluates the individual, the borrower is more indebted. Hence, the second bank will offer worse loan terms than the first. [Table A.VIII](#) mitigates this concern by controlling for the order in which the loan was granted. As shown, these results are very similar to our baseline.

Perk mechanism Equation 5 includes a complete set of interactions of district, bank and time fixed effects, as well as first-time borrower fixed effects and individual (more precisely, family member) fixed effects. However, one might wonder whether the results of [Figure 3a](#) are subject to more complicated sources of bias. For instance, an unobservable characteristic of some first-time borrower–family member pairs (e.g., a special affinity between them) may be correlated with both the bank selection of the first-time borrower and the credit behavior of the family member. We address this problem by exploiting the fact that a pair first-time borrower–family member appears more than once in our database if the family member is a client of more than two banks at the time credit is granted. We then work with an additional specification that adds fixed effects for every combination of first-time borrower and family member. One might also worry that some bank–customer relationships introduce a bias in our estimations. We exploit here the fact that a combination family member–bank may appear more than once in our database if the individual is related to more than one first-time borrower.³⁷ More specifically, we include fixed effects for every combination of family member and bank. [Figure A.3](#) in the Online Appendix presents the results. With this much stricter specification, the results remain similar to the original.

³⁷Notice that the dependent variable is the same if the first-time borrowers related to the same individual receive credit in the same period. However, the dependent variable does change if the first-time borrowers are included in the financial system in different years.

8 Conclusion

Scholars and policy makers are concerned to determine why access to credit is unequal. One possible explanation is that individuals who carry out informal activities do not have hard information to show banks, who are thus unable to use their usual scoring systems. The lack of access to first-credit greatly hinders the finance of these individuals throughout their lives because they cannot build a credit history. The problem is enormous in developing countries, but it also affects developed countries where some minorities engage in informal activities. Although it is known that an important group of banks, generally smaller and local, manage to serve these population segments in a limited fashion, little is known about the strategies that these banks use to substitute (or complement) the lack of official information.

In this paper, we use a large-scale dataset that contains the universe of loans in five regions of Peru to show that banks use family ties as a strategy to approach first-time borrowers. We leverage naming conventions in Peru to build family networks for every individual receiving a loan in each of his available banks. Exploiting variability in the relative importance of an individual's family across banks, we show that an individual has greater probability of receiving a loan in the bank where his family is better connected. We also provide empirical evidence suggesting that our results are partially driven by supply forces. For instance, we show that banks oriented towards individuals working in the formal sector do not use family networks to decide on whether to grant their loans. This is to be expected, since these banks will require individuals to provide documentation, such as income statements. Instead, we show that informal-oriented banks are the ones responsible of our results. Moreover, we show that the terms of these loans are better relative to those in other banks where his family is less important. With these results in hand, we describe two novel channels that may be driving informal-oriented banks to use family networks. Informal-oriented banks, pressed by competition, may find it convenient to grant a first loan to an existing client's family members. Banks may also assume that the individual being considered will have similar payment behavior than those of his family.

We believe our research opens new questions. For instance, do banks in other regions, including developed countries, also use family networks to substitute the lack of hard information of those

engaging in informal activities? If not, could they in the future? The fact that banks with characteristics similar to those we here call “informal-oriented banks” partially serve both the informal sector and minorities suggests this may indeed be the case. Another question is the potential of this strategy to address inequality within these other segments. In particular, if some families benefit and others do not, are inequalities somehow perpetuated? In general, we believe family networks hold great promise for the study of how these population segments access to credit. This paper is a step in that direction.

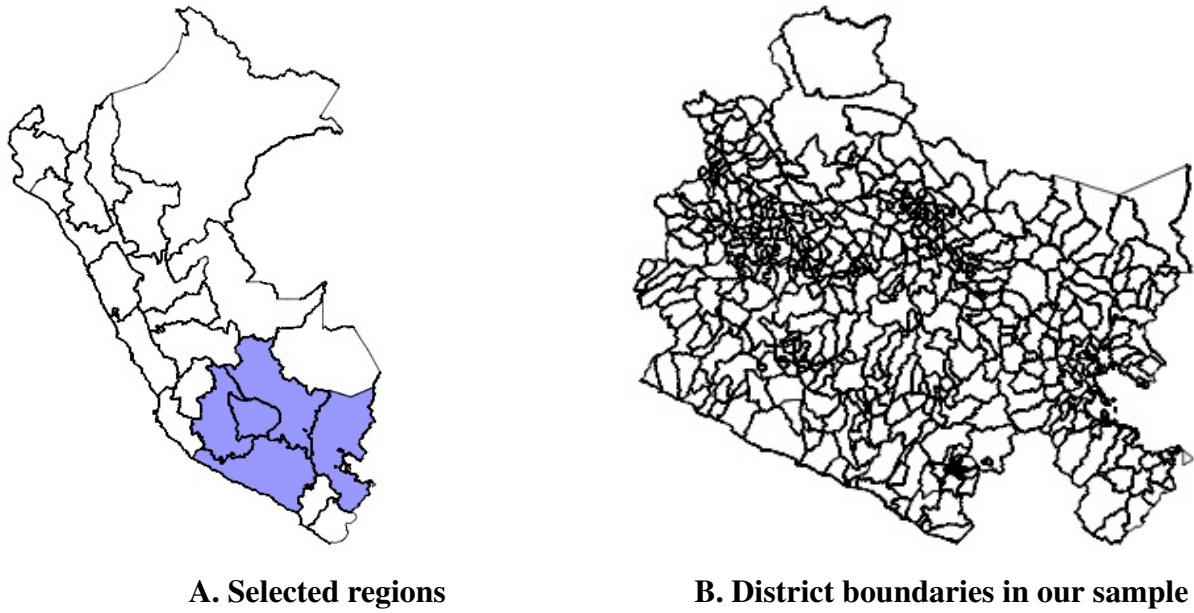
References

- Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2016). Network structure and the aggregation of information: Theory and evidence from indonesia. *American Economic Review* 106(7), 1663–1704.
- Alberti, G. and E. Mayer (1974). Reciprocidad e intercambio en los Andes peruanos. *Lima, Perú: Perú Problema* (12).
- Angelucci, M., G. De Giorgi, and I. Rasul (2018). Consumption and investment in resource pooling family networks. *The Economic Journal* 128(615), 2613–2651.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science* 341(6144).
- Barone, G. and S. Mocetti (2021). Intergenerational mobility in the very long run: Florence 1427-2011. *The Review of Economic Studies* 88(4), 1863–1891.
- Broecker, T. (1990). Credit-worthiness tests and interbank competition. *Econometrica: Journal of the Econometric Society*, 429–452.
- Carpio, M. A. and M. E. Guerrero (2021). Did the colonial mita cause a population collapse? What current surnames reveal in Peru. *The Journal of Economic History* 81(4), 1015–1051.
- Céspedes, N. (2017). La demanda de crédito a nivel de personas: RCC conoce a ENAHO.
- Cruz, C., J. Labonne, and P. Querubin (2017). Politician family networks and electoral outcomes: Evidence from the Philippines. *American Economic Review* 107(10), 3006–3037.
- Fafchamps, M. (2000). Ethnicity and credit in African manufacturing. *Journal of Development Economics* 61, 205–235.
- Fisman, R., D. Paravisini, and V. Vig (2017). Cultural proximity and loan outcomes. *American Economic Review* 107(2), 457–92.
- Gagliarducci, S. and M. Manacorda (2020). Politics in the family: nepotism and the hiring decisions of Italian firms. *American Economic Journal: Applied Economics* 12(2), 67–95.
- Garcilazo de la Vega, . (1609). *Comentarios reales de los Incas I*. Lisboa.
- Karlan, D. S. (2007). Social connections and group banking*. *The Economic Journal* 117(517), F52–F84.
- Loayza, N. V. (2016). Informality in the process of development and growth. *The World Economy* 39(12), 1856–1916.
- Marín, J., X. Barragán, and A. Zaballos (2014). Informe sobre la situación de conectividad de Internet y banda ancha en Perú. *SI: BID*.
- Murra, J. (1975). Formaciones económicas y políticas del mundo andino. IEP. *IEP, Lima*.

- Okten, C. and U. O. Osili (2004). Social networks and credit access in Indonesia. *World Development* 32(7), 1225–1246.
- Padgett, J. F. and P. D. McLean (2006). Organizational invention and elite transformation: The birth of partnership systems in renaissance Florence. *American Journal of Sociology* 111(5), 1463–1568.
- Wachtel, N. (2017). *Sociedad e ideología: ensayos de historia y antropología andinas*. Instituto de Estudios Peruanos.

ONLINE APPENDIX

Figure A.1: Selected regions and district boundaries



Note: Panel A shows the regions included in our sample, in blue. Panel B shows the district boundaries in the regions included in our sample.

Figure A.2: The effect of centrality on credit access, by number of top common excluded surnames

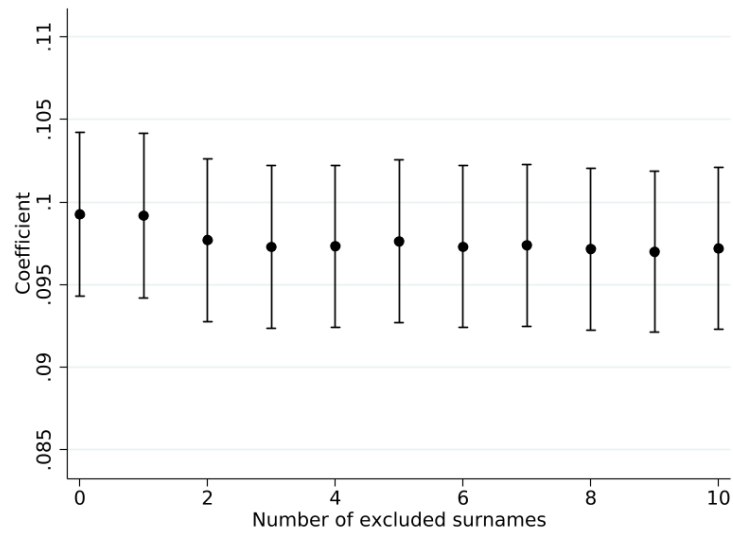


Figure A.3: The effect of granting credit to a first-time borrower on family members' behavior, using additional sets of fixed effects

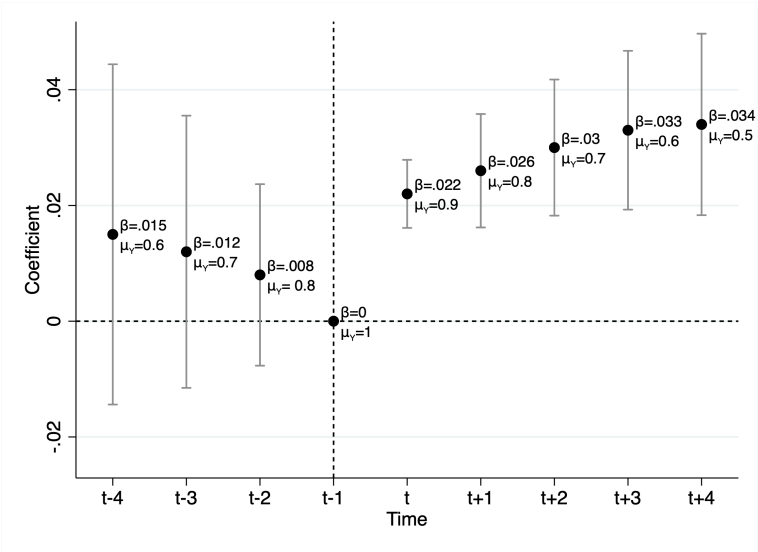


Table A.I: The effect of centrality on credit access, by different sets of fixed effects

	Credit Given									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Centrality [$t-1$]	0.7787*** (0.0303)	0.7774*** (0.0252)	0.3722*** (0.0131)	0.3655*** (0.0131)	0.5972*** (0.0205)	0.1702*** (0.0068)	0.1702*** (0.0068)	0.1340*** (0.0058)	0.1023*** (0.0049)	0.0993*** (0.0050)
Dist $\times$ Bank FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Dist $\times$ Time FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Bank $\times$ Time FE	No	No	No	No	No	No	No	Yes	Yes	Yes
Indiv FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Dist FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist $\times$ Bank $\times$ Time FE	No	No	No	No	No	No	No	No	Yes	Yes
Controls	No	No	No	No	No	No	No	No	No	Yes
Y Mean	2.53	2.53	2.53	2.53	2.53	2.53	2.53	2.53	2.53	2.53
N indiv	952,956	952,956	952,956	952,956	952,956	952,956	952,956	952,956	952,136	952,136
N	38,395,391	38,395,391	38,395,391	38,395,391	38,395,391	38,395,338	38,395,338	38,395,338	38,378,450	38,378,450

Note: This table shows the effect of family centrality on credit access for unbanked individuals, using different set of fixed effects. The dependent variable is a dummy indicating whether the individuals i , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. The explanatory variable is the eigenvector centrality. To simplify interpretation, it has been rescaled to have mean 0 and standard deviation 1. The fixed effects differ by columns. From columns (1) to (9) we add fixed effects one by one. Column (10) includes all the fixed effects plus control variables. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. Standard errors are in parenthesis. They are clustered at district $\times$ bank level. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.II: Descriptive statistics of first-time borrowers, by number of granting banks

Panel A. Population				
	Bancarized in 1 bank (1)	Bancarized in > 1 bank (2)	(1)-(2) Estimate	Tstat
% indiv with consumption credits	0.435	0.385	0.050***	31.890
% indiv with business credit	0.559	0.346	0.213***	137.408
% indiv with mortgage	0.003	0.000	0.003***	18.609
% of relatives with bad credit	20.199	19.869	0.330***	9.741

Note: The descriptive statistics were calculated for all the individuals of the sample. The sizes of the *Bancarized in one bank* and *Bancarized in more than one bank* samples are 842270 and 115161, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel B. ENAHO Sample				
	Bancarized in 1 bank (1)	Bancarized in > 1 bank (2)	(1)-(2) Estimate	Tstat
% in rural areas	0.577	0.447	0.130***	10.744
% female	0.496	0.484	0.012	0.990
% with high school or more	0.544	0.657	-0.114***	-9.306
% with with complete post HS education	0.175	0.232	-0.057***	-5.988
% main occupation is formal	0.171	0.240	-0.069***	-6.488
Avg. HH annual net income per capita	7,527.630	8,546.124	-1,018.494***	-4.764
% of relatives with bad credit	20.152	19.818	0.334	1.249

Note: The descriptive statistics were calculated for the individuals surveyed by the ENAHO between 2010 and 2019. The sizes of the *Bancarized in one bank* and *Bancarized in more than one bank* samples are 15440 and 1911, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.III: The effect of different centrality measures on credit access

	Credit Given			
	(1)	(2)	(3)	(4)
Eigenvector Centrality[t-1]	0.0993*** (0.0050)			
Degree Centrality[t-1]		0.0493*** (0.0025)		
Closeness Centrality[t-1]			0.1324*** (0.0059)	
Betweenness Centrality[t-1]				0.1050*** (0.0089)
Indiv FE	Yes	Yes	Yes	Yes
Dist $\times$ Bank $\times$ Time FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Y Mean	2.53	2.53	2.53	2.53
β /Mean	3.93	1.95	5.24	4.16
Y Mean	952,136	952,136	952,136	952,136
N indiv	38,378,450	38,378,450	38,378,450	38,378,450

Note: This table shows the effect of family centrality on credit access for unbanked individuals, using different measures of centrality. The dependent variable is a dummy indicating whether the individuals i_t , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. The explanatory variable differs by columns. Column (1) uses the eigenvector centrality. Column (2) uses the degree centrality. Column (3) uses the closeness centrality. Finally, column (4) used the betweenness centrality. To simplify interpretation, the four measures of centrality have been rescaled to have mean 0 and standard deviation 1. Standard errors are in parenthesis. They are clustered at district $\times$ bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.IV: The effect of centrality on credit access, using average centrality

	Regression					Stats		
	(1)	(2)		(3)	(4)			
	Centrality > 0	Centrality < Median centrality	Centrality > Median centrality	Continuous centrality	Continuous centrality [†]	N	Y Mean	N indiv
(1) Baseline	0.2999*** (0.0101)	0.1487*** (0.0104)	0.4884*** (0.0173)	0.0993*** (0.0050)	0.1748*** (0.0098)	38,378,450	2.5256	952,136
(2) Average Centrality	0.2999*** (0.0101)	0.1711*** (0.0107)	0.4413*** (0.0163)	0.0605*** (0.0036)	0.0880*** (0.0074)	38,378,450	2.5256	952,136
(3) Without indiv whose 2nd surname includes “De”	0.2978*** (0.0102)	0.1487*** (0.0105)	0.4835*** (0.0173)	0.0987*** (0.0050)	0.1733*** (0.0098)	37,272,427	2.5270	924,700

Note: This table shows the effect of family centrality on credit access for unbanked individuals, using different centrality measures. The dependent variable for all regressions is a dummy indicating whether the individuals i , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. The explanatory variable differs by columns. Column (1) uses a dummy equal to 1 when centrality is positive; otherwise its value is 0. Column (2) uses two indicator variables that use the median centrality for each individual conditional on this being positive. With this median, the variables are dummies indicating whether the centrality of a bank for an individual is above or below the median. Column (3) uses the continuous measure of eigenvector centrality. Finally, Column (4) uses the continuous measure of eigenvector centrality but restricting the sample to banks with number of relatives > 0. To simplify interpretation, the eigenvector centrality has been rescaled to have mean 0 and standard deviation 1. Standard errors are in parenthesis. They are clustered at district × bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.V: The effect of centrality on credit access, by different consumption sets

	Regression				Stats		
	(1)	(2)		(3)			
	Centrality > 0	Centrality < Median centrality	Centrality > Median centrality	Continuous centrality	N	Y Mean	N indiv
(1) Baseline	0.2999*** (0.0101)	0.1487*** (0.0104)	0.4884*** (0.0173)	0.0993*** (0.0050)	38,378,450	2.5256	952,136
(2) Same bank type of bank of first given credit	0.3440*** (0.0132)	0.2217*** (0.0140)	0.4903*** (0.0179)	0.0906*** (0.0053)	25,668,350	3.7638	949,374
(3) Banks with at least 0.5% mkt share	0.6110*** (0.0204)	0.4218*** (0.0196)	0.8691*** (0.0284)	0.3027*** (0.0131)	19,011,907	4.8846	916,504
(4) Banks with at least 1% mkt share	0.8396*** (0.0284)	0.6134*** (0.0272)	1.1253*** (0.0370)	0.4333*** (0.0193)	14,647,928	6.0624	880,422
(5) Banks with at least 2% mkt share	1.3684*** (0.0481)	1.0459*** (0.0457)	1.7281*** (0.0594)	0.7313*** (0.0349)	9,588,919	8.4083	807,113
(6) Banks with at least 3% mkt share	1.8789*** (0.0684)	1.4400*** (0.0653)	2.3242*** (0.0830)	1.0285*** (0.0541)	6,874,267	10.7346	744,356
(7) Banks with office in district	0.8186*** (0.0819)	0.5936*** (0.0729)	1.2303*** (0.1065)	0.7237*** (0.0937)	4,546,023	7.7986	351,728

Note: This table shows the effect of family centrality on credit access for unbanked individuals, using different consumption sets. The dependent variable for all regressions is a dummy indicating whether the individuals i , resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. The explanatory variable differs by columns. Column (1) uses a dummy equal to 1 when centrality is positive; otherwise its value is 0. Column (2) uses two indicator variables that use the median centrality for each individual conditional on this being positive. With this median, the variables are dummies indicating whether the centrality of a bank for an individual is above or below the median. Finally, Column (3) uses the continuous measure of eigenvector centrality. To simplify interpretation, the eigenvector centrality has been rescaled to have mean 0 and standard deviation 1. Standard errors are in parenthesis. They are clustered at district×bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. $p < 0.10$, $** p < 0.05$, $*** p < 0.01$

Table A.VI: The effect of centrality on credit access, by family reputation: alternative measure

	Credit Given						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Centrality > 0, [t-1]	0.2999*** (0.0101)	0.2980*** (0.0104)	0.3236*** (0.0107)				
Centrality > 0, [t-1] × D-B-I: Above median rel bad credit[t-1]		0.0074 (0.0115)					
Centrality > 0, [t-1] × D-B-I: Above 25% Ratio rel bad credit[t-1]			-0.0939*** (0.0141)				
Centrality [t-1]				0.0993*** (0.0050)	0.1108*** (0.0056)	0.1043*** (0.0050)	0.1080*** (0.0057)
Centrality [t-1] × D-B-I: Ratio rel bad credit					-0.0007*** (0.0001)		
Centrality [t-1] × D-B-I: Above median rel bad credit[t-1]						-0.0121 (0.0075)	
Centrality [t-1] × D-B-I: Above 25% Ratio rel bad credit[t-1]							-0.0391*** (0.0084)
Indiv FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist × Bank × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	2.53	2.53	2.53	2.53	2.53	2.53	2.53
N indiv	952,136	952,136	952,136	952,136	952,136	952,136	952,136
N	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450	38,378,450

Note: This table shows the effect of family centrality on credit access by family's reputation. The sample is composed of unbanked individuals. The dependent variable is a dummy indicating whether the individual i resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. Each column presents a different regression. Column (1) uses only a dummy equal to 1 when centrality is positive. Column (2) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (3) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Column (4) uses only continuous centrality. Column (5) uses the ratio of relatives with bad reputation. If the individual has 0 relatives the ratio is 0. Column (6) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (7) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Standard errors in parenthesis. They are clustered at district×bank level. We control for the number of relatives in the network in $t-1$, the relatives' credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$ * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.VII: The effect of centrality on credit access, by family reputation: restricted sample

	Chosen bank						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Continuous centrality [t-1]	0.1748*** (0.0098)	0.1838*** (0.0106)	0.1827*** (0.0106)	0.1854*** (0.0105)	0.1832*** (0.0104)	0.1667*** (0.0100)	0.1838*** (0.0105)
Continuous centrality [t-1] × D-B-I: Ratio rel w/ bad reputation		-0.0005** (0.0002)					
Continuous centrality [t-1] × D-B-I: Above median rel w/ bad reputation [t-1]			-0.0153 (0.0098)				
Continuous centrality [t-1] × D-B-I: Above 25% rel w/ bad reputation [t-1]				-0.0381*** (0.0109)			
Continuous centrality [t-1] × D-B-I: Ratio rel bad credit					-0.0005*** (0.0002)		
Continuous centrality [t-1] × D-B-I: Above median rel bad credit [t-1]						0.0162* (0.0096)	
Continuous centrality [t-1] × D-B-I: Above 25% rel bad credit [t-1]							-0.0344*** (0.0108)
Indiv FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist × Bank × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Y Mean	4.54	4.54	4.54	4.54	4.54	4.54	4.54
N indiv	919,150	919,150	919,150	919,150	919,150	919,150	919,150
N	18,168,583	18,168,583	18,168,583	18,168,583	18,168,583	18,168,583	18,168,583

Note: This table shows the effect of family centrality on credit access by family's reputation. The sample is composed of unbanked individuals and restricted to banks with number of relatives >0. The dependent variable is a dummy indicating whether the individual i resident of district d and recipient of the first loan in period t , receives credit from bank b . For ease of comparison, this variable has been multiplied by 100. Each column presents a different regression. Column (1) uses a dummy equal to 1 only when centrality is positive. Column (2) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (3) uses an interaction between positive centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Column (4) uses only continuous centrality. Column (5) uses the ratio of relatives with bad reputation. If the individual has 0 relatives the ratio is 0. Column (6) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit is larger than its median. Column (7) uses an interaction between continuous centrality and a dummy equal to 1 when the share of family members with bad credit of the individual is larger than 25%. Standard errors in parenthesis. They are clustered at district × bank level. We control for the number of relatives in the network in $t-1$, the relatives' credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.VIII: The effect of centrality on credit conditions

	Log (credit) (1)	Interest rate (PEN) (2)	Credit quality (6m later) (3)
Panel A: All Districts			
Centrality [t-1]	0.0134 (0.0084)	-1.5100** (0.7027)	0.0067 (0.0042)
Y mean	7.1568	48.5109	0.7971
N indiv.	82,472	57,757	57,367
N obs.	178,635	123,532	123,724
Panel B: Districts with larger % of formal-oriented banks			
Centrality [t-1]	0.0031 (0.0124)	-0.3920 (0.9983)	0.0068 (0.0074)
Y mean	7.1744	47.8327	0.8010
N indiv.	31,098	21,794	20,134
N obs.	66,312	45,892	42,860
Panel C: Districts with smaller % of formal-oriented banks			
Centrality [t-1]	0.0223* (0.0122)	-2.8759*** (1.0135)	0.0063 (0.0049)
Y mean	7.1432	48.9580	0.7943
N indiv.	50,512	35,234	36,546
N obs.	110,363	76,022	79,313
Indiv FE	Yes	Yes	Yes
DistXBankXTime FE	Yes	Yes	Yes
First credit control	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: This table shows the effect of family centrality on credits conditions. All regressions have individual fixed effects, District×Bank×Time fixed effects and controls. Additionally, all regressions add first credit control that equals 1 if it is the first institution with credit given. The sample used differs by panel. The first panel considers individuals that got their first loans in more than one bank within a three month window and the banks in which they got their first loans. The second and third panel use subsamples. The second panel includes only the individuals that live in districts with larger share of formal-oriented institutions available than the median district. The third panel includes those living in districts with smaller share of formal-oriented institutions available than the median district. The explanatory variable is eigenvector centrality (rescaled to have mean 0 and standard deviation 1). The dependent variable differs by column. Column (1) uses the logarithm of the loan amount. Column (2) uses the interest rate of loans made in the Peruvian currency (PEN). Column (3) uses a dummy equal to 1 if the loan has a good credit quality six months later. Standard errors are in parenthesis. They are clustered at district×bank level. We control for the number of relatives in the network in $t-1$, the relative's credit size in $t-1$ and the number of relatives in the network with good credit quality in $t-1$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.