



Microlending on mobile social credit platforms: an exploratory study using Philippine loan contracts

Jian Mou¹ · J. Christopher Westland² · Tuan Q. Phan³ · Tianhui Tan³

Published online: 25 November 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Microlending has grown rapidly and now benefits around 250 million people globally, half who would otherwise not have access to credit. Use of social credit systems for microlending risk assessment is most pronounced in Asia, as most Western countries tightly regulate personal information available to lenders. In most of the developing world, geography, social structure, disease, climate and culture have a much stronger influence on credit risk and borrowing than do governmental and corporate systems. In this study, we obtained 784 loan contracts with 3577,912 personal communications and locations. Exploratory analysis found loan default depends on social network structure; graph analysis indicated that those who were likely to default tended to communicate with other likely defaulters. Detailed tests were equivocal, suggesting that social network communication structure provided little additional information to predict default, and may even add noise to the data. Our tests strongly supported the importance of location and proximity to particular sorts of landmarks on the potential for default. Proximity to some landmarks, e.g. city hall, moving companies and train stations, were associated with lower loan default. Others, such as parks, stadiums and bus stations, were correlated with a higher loan default. We restructured our tests based on risk-return versus loan default effect with little change in results.

Keywords Social credit systems · Social networks · Privacy · Credit scoring · Microlending · Peer-to-peer lending · Credit risk

1 Introduction

Social credit systems are countrywide reputation systems intended to standardize the assessment of citizens' and business' economic and social reputation, or 'credit'. Sachs [56] argues that in most of the developing world, geography, social structure,

✉ Jian Mou
jian.mou@outlook.com

Extended author information available on the last page of the article

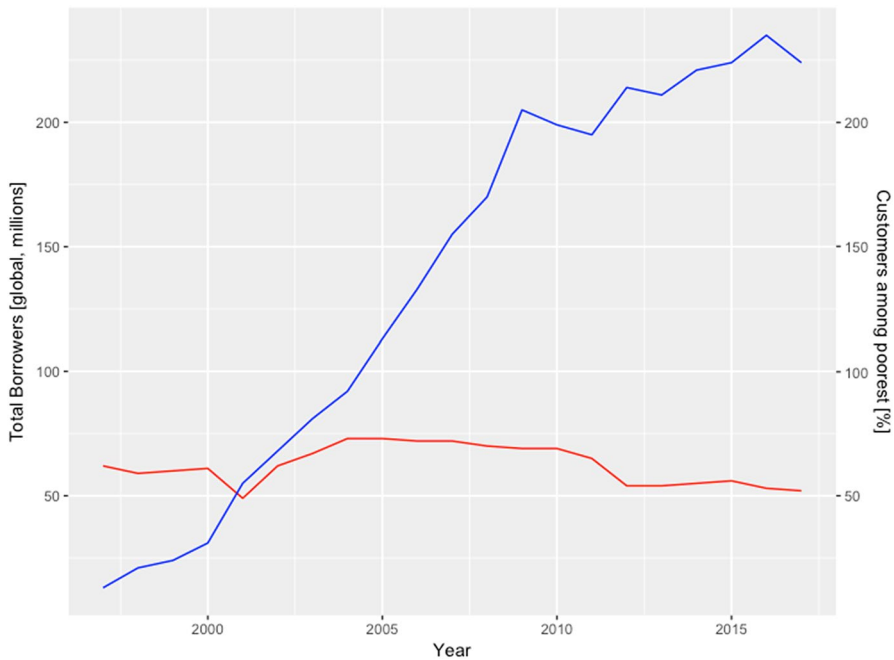


Fig. 1 Number of microloans globally and social welfare of the poor. Adapted from Cull et al. [17]

disease, climate and culture have a much stronger influence on income, credit risk and borrowing than do governmental and corporate systems. This makes traditional lending nearly impossible because of the difficulty in determining default risk. This problem has been detailed in Sanchez et al. [59] and McCord and Sachs [44] in their country specific studies.

Our current research investigates the following research question: “Is data and GPS tracking from mobile social networks useful in reducing the risk-related loan costs (including default) in microlending?” Our research question is important for two reasons: (1) microlending has grown rapidly and now benefits around 250 million people globally, half who would otherwise not have access to credit (Fig. 1); and (2) risk assessment and control of loan defaults is the major cost of underwriting these loans and is passed on to the borrowers in very high interest rates. One of the principal challenges of microfinance is providing small loans at an affordable cost. Loan providers have argued that if they were able to obtain more ‘personal’ or ‘social’ information germane to credit, those costs would be lower because loan default risk would be lower [17, 48]. Mimouni [47] found the global microfinance interest rate to average 35%, varying between 17% and 80%. The reason for high interest rates is not primarily cost of capital, as many local microfinance organizations receive zero-interest loan capital from the online microlending platforms like Kiva. Rather, the main reason for the high cost of microfinance loans is the high transaction cost of traditional microfinance operations relative to loan size, with most of this dedicated to predicting the risk of default [17, 48]. Figure 1 shows that

despite high interest and fees, microlending is expanding rapidly, opening access to capital previously unavailable to many of the world's poor.

Our current research seeks to advance our understanding of our research question through the conduct of an empirical study using proprietary information from a Manila, Philippines financial firm of 784 microloans with an associated 4 million social network data points from SMS/voice communications with GPS location data. We argue that our specific dataset will allow us to draw conclusions about a simple but representative social credit system for predicting default based on social network data. Our research credit scoring will limit analysis to two types of social information: (1) network structure of communications with the borrower, which refers to using graph statistics to investigate the network structure of loans and borrowers' communications, and (2) geographical location of the borrower. These two nonetheless are important predictors in nearly all social credit systems, though in full-scale systems such as China's social credit system there exist numerous other demographic and personal predictors beyond the scope of this study.

The research proceeds as follows. Section 2 reviews the relevant literature on microlending, social graph metrics, credit scoring, and media synchronicity theory. Section 3 describes the dataset, its source and exploratory statistics. Section 4 tests the research hypotheses for loan default and profitability; and Sect. 5 summarizes our conclusions and discusses the implications for microlending and future research.

2 Literature review and hypothesis development

2.1 Microlending

New mobile and cloud based technologies have provided alternatives for IT based social credit systems for microlending [27, 32, 58, 62, 65, 69, 77, 78]. The most ambitious and controversial project thus far has been China's Social Credit System [46] that bases access to credit on behavior, associations, social history, and location. Though social credit scoring has been met with skepticism or hostility by privacy groups, such systems have found widespread application over the past 5 years in developing economies, allowing the granting of loans to those who previously would have found access to capital impossible. De Soto [21, 22] has cited access to capital as a major impediment to improving the lot of the poor; thus, social credit systems can be an important step in unlocking the potential of a countries' individuals and small businesses. Additionally, the poor in Asia favor access to credit more than they do privacy—especially in villages where little privacy may be had to begin with [30].

Microlending is generally considered to have originated with the Grameen Bank founded in Bangladesh in 1983 [74–76]. Since then, even traditional banks have introduced microlending and, today, it is widely used in developing countries as a tool for poverty alleviation among rural and female citizens. Much of the controversy over microlending stems from difficulties in assessing credit risk of potential recipients, who may have had few financial transactions of record in their past.

Alternatives for assessing credit risk (what is called ‘credit scoring’) may emphasize social credit data in microlending.

Aside from privacy concerns, there remain issues concerning just how effective social credit systems are in controlling lending risk. Even though lenders may have more information, they may not be able to use this to more effectively assess credit risk. Failures in developing microlending have been studied widely. For example, Riggins and Weber [55] emphasized two types of problems which may inhibit the success of microlending: (1) information asymmetries where the lender has inadequate loan borrower information; and (2) identification bias may result in loans being granted for reasons other than the potential success of the business venture. Further, Shi [63] pointed out that there are many P2P platforms that failed, due to cash shortage, small capitalization, poor management or fraud, and were no longer able to continue to service clients. Moreover, many loan projects fail to reach the funding goals, and such failures reduce the likelihood of success of subsequent crowdfunding projects [39]. Therefore, Yan et al. [73] argue that improving the success rate of fund formation and contributing to the sustainability of social lending has become an important issue in the context.

2.2 Social network research

At the micro-level, social network research typically begins with an individual, snowballing as social relationships are traced, or may begin with a small group of individuals in a particular social context [60, 61, 68]. Micro level analysis may occur at the level of the dyad a social relationship between two individuals where the focus analysis is the structure of the pairwise relationship and tendencies toward reciprocity/mutuality; or the triad where they concentrate on balance and transitivity [35]. Studies that focus on the individual in a social setting are called “ego networks,” where the focus is individual nodes or “actors” in the network [4, 26, 40]. They focus on characteristics such as size, relationship strength, density, centrality, prestige and roles such as isolates, liaisons, and bridges [34] and are used in the fields of psychology, ethnography and genealogy [19].

At the macro-level, analyses trace the outcomes of interactions, such as economic resource transfers among a population. Macro-level social networks display features of social complexity, which involves substantial non-trivial features of network topology, with patterns of complex connections between elements that are neither purely regular nor purely random and are distinguished by a heavy tailed degree distribution, a high clustering coefficient and community structure [8, 10, 24, 64].

Privacy regulations, e.g., which guarantee the ability of an individual or group to hide information about themselves, insert opportunities for moral hazard an adverse selection into the microlending business. Böhme and Pöttsch [7] asserted that the financial and social objectives of lending are inherently incompatible; in a case study, Grodzinsky and Tavani [29] revealed exactly how those objectives might diverge; while Dillon and Lending [23] analyzed the reduced accuracy that accompanies improvements in privacy. In most developed economies of the world, financial privacy is rigorously protected by government [37, 66].

Lenders often complain that were they only able to know more about their lenders' daily activities, communications, and so forth, that they could reduce the cost of lending, offer their customers better rates, and guarantee their investors more profitability [55].

We considered media synchronicity theory [20] a loose, interpretive, and conjectural dialectic on media richness [18]. Media synchronicity is neither predictive, nor is it particularly descriptive in a scientific sense; it is also difficult to model except in the most ad hoc sense. Though suggestions made that this might be productive, it was generally found to be of little relevance to the current analysis.

Lending focus is typically dominated by earnings, and the models typically are credit scoring models, with particular predictors related to prior behavior, finances and lifestyle. We extend those predictors here to include communications on mobile platforms. There are many modes of communications, with an evolving landscape that increasingly favors visual communications. The two streams of communication used here are textual, and probably the major text communications channels used on mobile platforms for the particular geographical location of the system provider.

We determined the likelihood of loan default using:

1. "baseline" credit scoring contained in individual loan contract information,
2. additional "social" information summarized in the graph structure of communication links in the borrowers' social network,
3. additional "location" information obtained from GPS data.

Based on above theoretical backgrounds, we tested five nested hypotheses concerning the binary factor "loan default" derived from our research question:

H_{1a} (baseline): Loan default is predicted by loan interest rate and principal.

H_{1b} (baseline): Loan default is predicted by loan duration, interest rate and principal.

H_{2a} (social network 1): Loan default is predicted by whether communications are predominantly ingoing or outgoing SMS or voice communications

H_{2b} (social network 2): Loan default is predicted by social network structure as captured in centrality metrics of the message (either phone or SMS) sender versus receiver.

H_{3a} (location): Loan default is predicted by proximity to location-specific business and organizational landmarks. We established a critical value of 50 ms proximity of a landmark identified on Google Maps to the point of a call or SMS message location

H_{3b} (loan contract information + location): Loan default is predicted by all of the information in H_1 and H_{3a} .

H_{3c} (loan contract information + social network + location): Loan default is predicted by all of the information in H_1, H_{2a}, H_{2b} and H_{3a} .

Nesting our hypotheses allows us to assess the marginal value of loan default information as it is added to our representative "social credit system."

3 Data and network structure of communications with the borrower

We used Manila, Philippines data in our analysis, where privacy laws are relatively lax, and the microlending company is allowed access to complete phone records of borrowers under contract. The dataset was limited to the data that could actually be acquired from the company under non-disclosure. The data set combines short message services (SMS) communications with voice communications for each handset. SMS communications are by default given an average communications duration of one minute in this research, which was an estimate of the equivalent amount of information that would be communicated by voice. Call duration was interpreted as a surrogate for information conveyed in a communication, with more information indicating a “stronger” relationship between a caller and receiver. We analyzed 1.5 years of call data from January 1, 2014 to June 27, 2015 with loan origination dates running from 2012-12-15 to 2014-08-11. Calls originated between 0 days after the loan origination to 366 days after the loan origination. In our dataset, 19.39% of the 784 loans were in default, representing 21.73% of loaned value (\$2346,040/\$10,795,722) (see Table 1).

Our raw data contained 4142,474 individual SMS and voice communications, or around ~5300 calls per borrower per year. We curated these 4142,474 individual communications, removing informational or emergency calls, standardizing all caller identifiers into 11-digit telephone numbers used in the country, and eliminating error and nonsense identifiers. The resulting 3577,912 caller identifiers were aggregated on caller-receiver dyads summing over the call durations to generate an edge-link list where links identify the quantity of information exchanged between individuals.

Prior research by Barabási [5, 6], Ebel et al. [25], Onnela et al. [50, 51] and Xia et al. [71] has suggested that that telephone email and SMS communications may be modeled as scale-free networks driven by underlying preferential attachment mechanisms. If this is the case for our dataset, then we might use this information to define a “structural model” for the exchange of information by borrowers in the empirical setting. Lucas [43] championed the idea of first creating a policy-invariant structural model of the underlying processes being investigated prior to model fitting. He argued that empirical studies too often assumed simplified models—particularly linear models—that without assumptions about the structure of cause and effect, research would tend to over-fit data. Scale-free networks were initially studied in networks of citations between scientific papers by Price [53]. Scale-free networks tend to have a few vertices with a degree that greatly exceeds the average—called “hubs” [6, 10, 12].

We began our exploratory study of the network structure of loans and borrowers’ communications with graph statistics. Modularity measures the strength of division of a network into communities and lies in the range $[-1/2, 1]$ and is the

Table 1 Research dataset

Dataset	Count	Descriptors
Loans	784	"Borrower"; "default"; "amount"; "interest"; "contracting time"
Voice communications by borrowers	564,562	"Borrower"; " call to/from"; "type"; "call time"; "duration"
SMS Communications by borrowers	3577,912	"Borrower"; " SMS to/from"; "type"; "SMS time"; "duration = 1"
Businesses within a 50-meter radius at time of communication by borrowers (obtained by comparing latitude–longitude with businesses listed on Google maps)	81,050	"Borrower"; "longitude"; "Latitude"; "call time"; "Time since loan contract"; ["set of 107 Google Maps business types (counts in 50 m radius of phone)"]
Communication graph size (vertices) after aggregation on duration	112,487	"Borrower"; "default (blue = default)"; "information" (edge weight)"

Table 2 Graph statistics for graph of mobile and SMS communications

Nodes	112,487
Edges	121,392
Zero deg nodes	0
Zero InDeg nodes	406
Zero OutDeg nodes	111,775
Non-zero in–out deg nodes	306

fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random. Our research graph has a relatively low modularity of 0.246 suggesting that link structure is not random, but it lacks strong community structure; this was confirmed by an analysis of spectral clustering on the graph Laplacian which shows sparse clustering.

As would be expected, most calls being made in this database are from the borrower to somebody outside of our loan selection (the “Out Degree” nodes in Table 2) because the company only provided data for a small subset of loans they extended in this geographical area.

Further analysis found that within subsets of loans, high risk borrowers (those who defaulted) are more likely to link exclusively to other high-risk borrowers (blue nodes in Fig. 2). Defaulters appear mainly to talk with other defaulters, while borrowers who repay loans appear to talk with other borrowers who repay their loans. Figure 2 shows representative samples at different scales of this phenomenon, and confirms a major preconception motivating social credit systems—that there is loan risk/default information contained in the social network that is not available in the basic demographics and history of the borrowers.

Figure 3 shows the same graphs but with the *vertices* sized by degrees (the number of outgoing communications) and *edges* sized by the total communication duration.

A maximum likelihood fit of the degree distribution of the resulting graph follows a Pareto distribution with very close fit and at scales that vary with domain. The cumulative probability distribution (Pareto) of our empirical graph’s degree distribution is:

$$F(x) = \begin{cases} 1 - \left(\frac{scale}{x}\right)^{shape} & \text{for } x \geq scale \\ 0 & \text{elsewhere} \end{cases}$$

The empirical values for this degree distribution under varying random subsets of the loans is presented in Table 3.

Table 3 shows that there are substantial disruptions in connectivity when you split the dataset into good and bad borrowers. The Akaike Information Criterion metrics (AIC) are generally good, and parameters are generally stable from left to right as we remove data from the MLE fitting. The AIC penalizes the number of free parameters, and lower values of AIC are preferred [33]. The average distance between nodes in this network is small compared to highly ordered networks such as a lattices [12]. Diameter drops significantly when there is a major perturbation

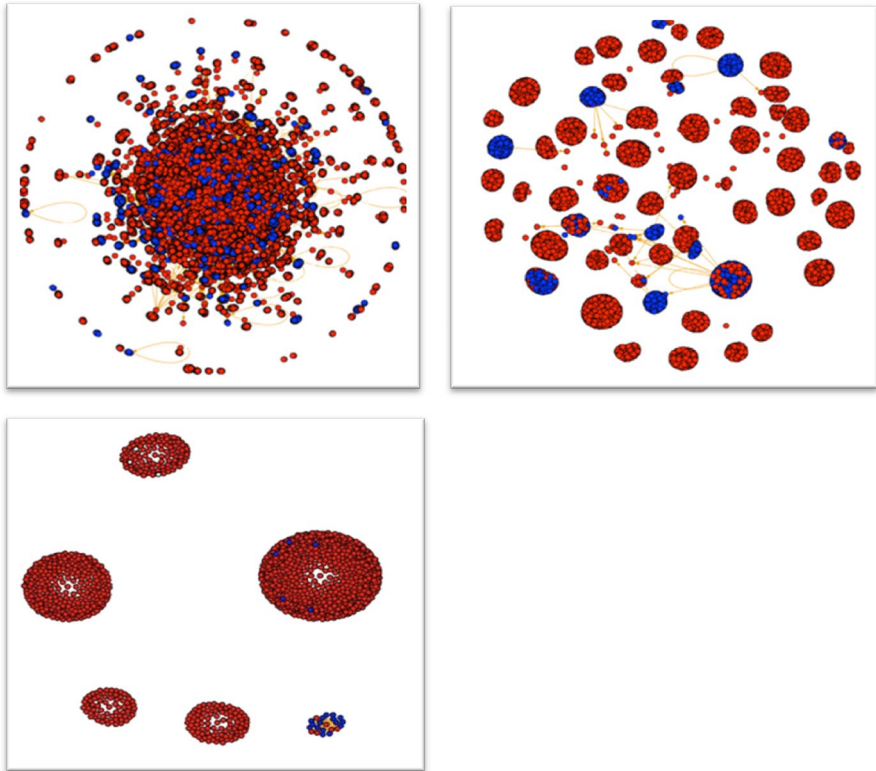


Fig. 2 Force-direct Graphs of communications graph structure for borrowers for all 784 loans (top); for a random subset of 70 loans (center); and for a random subset of 7 loans (bottom) using Fruchterman–Reingold layouts, where force is weighted by total call duration for a particular dyad; Blue colored nodes represent a default. (Color figure online)

caused by fragmentation of “hubs” from connectedness with the rest of the communication network. These are features we would expect to see in a scale-free network, as depicted in Fig. 4 plot of the degree distribution,

We examined network structure and the previously stated confounding effects using the full dataset of $784 \times 3,577,912$ semi-joins of loans and communications. The dataset was used to create directed graphs wherever $n > 2$ communications occurred, edges were defined for the $\binom{n}{2} = \frac{n!}{2(n-2)!}$ combinations of nodes.¹ The comprehensive graph was used to analyze the empirical graph structure and the influence of an individual borrower using whole graph metrics. Default risks were analyzed using general linear models of communication, location and loan

¹ We have accomplished graph construction data cleaning and plotting using the *igraph* package in R ([15]; [16]). Graph statistics for centralization, degree, modularity, community structure, diameter, connectedness and other statistics were computed using SNAP [41].

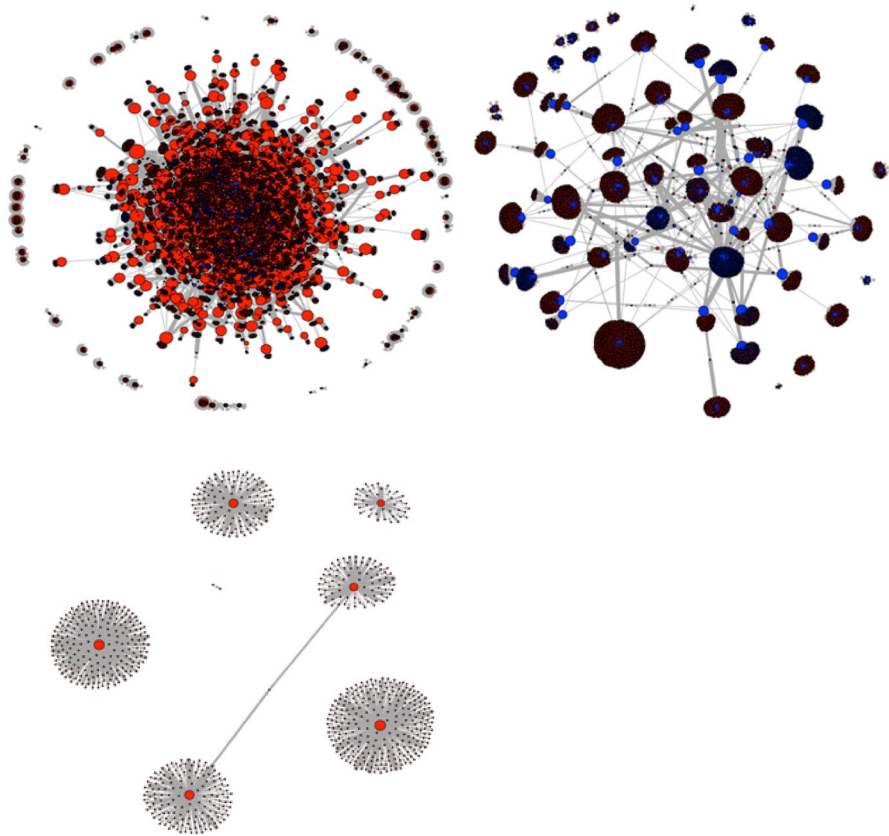


Fig. 3 Force-directed graphs with vertex sized by degree, and edges sized by total duration of communication for all 784 loans (top); for a random subset of 70 loans (center); and for a random subset of 7 loans (bottom) using Fruchterman–Reingold layouts, where force is weighted by total call duration for a particular dyad; Blue colored nodes represent a default; red colored nodes tend to be washed out by the gray edges

Table 3 Degree distribution (Pareto-discrete) of borrowers’ communications

	Full dataset		10% random sample (mean of 10 trials)		1% random sample (mean of 10 trials)		Good borrowers		Bad borrowers	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Shape	3.52	0.46	2.73	0.23	4.07	0.62	6.03	0.06	5.96	0.12
Scale	0.01	0.00	0.00	0.00	0.02	0.00	6.57	0.07	6.44	0.15
AIC	−3290.00		−5590.00		−2317.00		233,326.00		52,306.00	
Diameter	4.00		1.00		1.00		3.00		2.00	

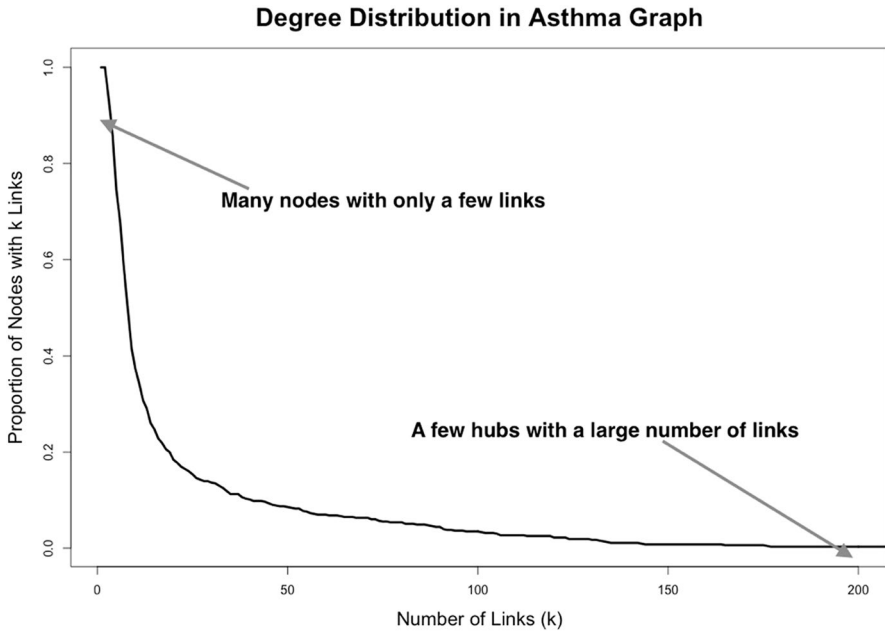


Fig. 4 Degree distribution of communications (shape = 3.5217)

Table 4 default ~ graph statistics regression

	Estimate	Std.	t	Pr(> t)
(Intercept)	-0.04097	0.00410	-10.00500	0.00000***
Number of out edges	-0.07558	0.00331	-22.83200	0.00000***
Number of in edges	-0.09644	0.00196	-49.34000	0.00000***
The number of triads that the node participates in	0.00262	0.00005	54.11900	0.00000***
Total communication duration	0.00000	0.00000	-0.06800	0.94600
Amount of loan	0.00001	0.00000	73.67600	0.00000***
Loan interest rate	0.03287	0.00062	52.85600	0.00000***
Eigenvector centrality	8489.00000	372.40000	22.79300	0.00000***
Farness centrality	0.00000	0.00000	1.32900	0.18400

R^2 0.09585, F-statistic 1610 on 8 and 121,383 DF

contract variables using appropriate residual distributions and link tables. We regressed these against an effect indicator variable for default, in Table 4. The value $R^2 \cong 9.6\%$ is relatively low, though not unusual for credit scoring algorithms where any additional variance explained can result in additional profitability. Investment models, credit scoring models and racetrack betting models often have fits where R^2 is less than 10% and in some cases lower than 1% (e.g., [38,

31, 67]. Over repeated use, these small amounts eventually add to increased profitability.

Regression statistics confirm earlier analysis of clustering of good and bad borrowers (e.g., in graphs in Figs. 1 and 2). Good borrowers tend to be better connected to the rest of the graph than are borrowers who are likely to default. As the number of *out* edges increases for the average borrower node, the default rate drops by 7.6% for each additional out edge; for each additional *in* edge it drops by 9.6%. The average borrower node participates in 4.5 triads but increasing this participation rate by a single triad only increases default rate by 1/5 of 1%. The loan amount and the loan interest rate are strongly correlated to default, which makes sense because credit scoring is going to allocate smaller loans and higher interest rates to borrowers who are likely to default.

4 Hypothesis tests

We conducted our tests of nested hypotheses using logistic regression models; specifically we applied a generalized linear model from the binomial family using a logit link function where “loan default” is the binary dependent factor and we assume continuous predictors and maximum likelihood estimation. In this form, fit scores are typically information measures, and we chose to assess our models with the Akaike Information Criterion metric [2, 3].

4.1 Tests of baseline model

H_{1a} (baseline): Loan default is predicted by loan interest rate and principal.

Term	Estimate	Std. error	Statistic	<i>p</i> value
(Intercept)	− 3.8812	0.0070	− 557.1217	0.0000
Principal amount	0.0001	0.0000	302.4569	0.0000
Interest rate	0.2654	0.0009	283.2373	0.0000
AIC	2300,451			

H_{1b} (baseline): Loan default is predicted by loan duration, interest rate and principal.

Term	Estimate	Std. error	Statistic	<i>p</i> value
(Intercept)	− 0.4728	0.0525	− 8.9977	0.0000
Principal amount	0.0001	0.0000	74.9205	0.0000
Interest Rate	− 0.0697	0.0054	− 12.9900	0.0000
Loan Duration	− 0.0047	0.0001	− 53.7488	0.0000
AIC	2248,829			

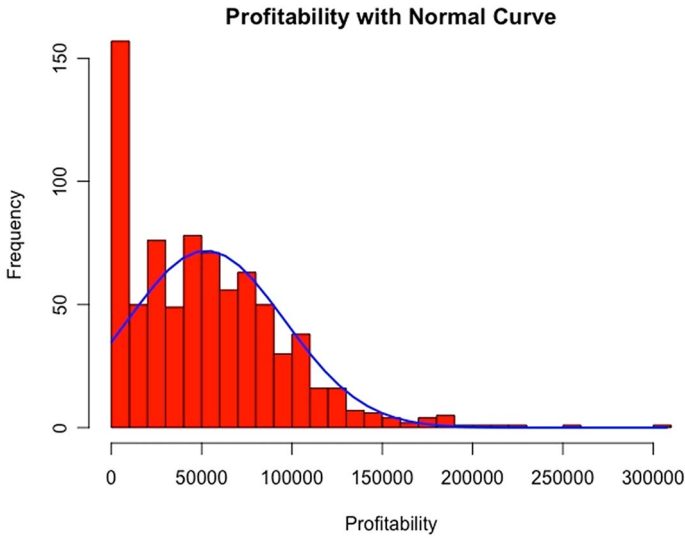


Fig. 5 Distribution of synthetic profitability with a normal density curve fitted to that data

Credit scoring algorithms at their most basic are “classifiers”—they dichotomize borrowers into “re-payers” and “defaulters”. More sophisticated algorithms provide feedback to determine interest rate and maximum credit, recognizing that loans can be restructured, borrowers may default on the entire principal or only a portion of it, or only miss an interest payment. Loan profitability derives from several factors—annual loan revenue is interest multiplied by principal; expenses derived from defaults, where the actual outcome of a default is complex. Once the borrower is in default the loan may be rescheduled, or the collateral may be sold and used to offset the loss from the loan default.

In practice, lenders are willing to take on riskier borrowers in exchange for collateral or at higher interest rates; this is what the finance community calls a risk-return trade-off. We looked at the potential for a more nuanced set of lending criteria that could take this into account, and constructed the following surrogate return rather than a binary ‘default’ risk:

$$\pi = \text{synthetic annual profitability} = \begin{cases} \text{principal} \times \text{interest} & \text{if good loan} \\ 0 & \text{if default} \end{cases}$$

Our synthetic profitability contains all the information in principal, interest and defaults to provide the best decision metric for profitability using our predictors on the database. Figure 5 graphs the histogram of synthetic profitability for the 784 loans on the dataset.

Our synthetic profitability metric is zero inflated with approximately 19.4% of loans having a value of zero. We fit the data to a zero inflated Poisson distribution (Tables 5, 6) using the Vuong test to compare to ordinary least squares with

Table 5 Test of zero-inflated poisson distribution (model = profit ~ 1)

	Estimate	S.E.
Count model coefficients (Poisson with log link)	6.47732	0.00156
Zero-inflation model coefficients (binomial with logit link)	-1.42501	0.09034
Log-likelihood	-71,870 on 2 df	
Vuong non-nested hypothesis z-statistic	13.97480; <i>p</i> value = 0.0000; model is indistinguishable from GLM Poisson with log link	

Table 6 Test of Zero-Inflated Tobit Distribution (model = profit ~ 1)

	Estimate	Std.	z	Pr(> z)
(Intercept)	46,940	1912.00	24.55	0.0000***
Log(scale)	11	0.03	367.73	0.0000***
Log-likelihood	-7919			
Vuong non-nested hypothesis z-statistic	15.390; <i>p</i> value = 0.0000; model is indistinguishable from OLS Normal			

normal errors, and concluding that the zero inflation is small enough at 19.4% that we can safely use ordinary least squares regression for our analysis.

Timing of communications appeared not to have an impact on profitability. Profitability is what we are using in the model rather than default. But just as with loan default, we might expect more defaults from people that are up all night (e.g., call at 3 am vs. 3 pm). The *diff_day* predictor (mean = 212.3) is the difference between the time of communication and the time of the original loan contract. It appears that loan profitability is not heavily dependent on time or the timing of communications, as the estimator value of -3.5 suggests that the influence of time never contributes or subtracts more than about \$1000 dollars from profitability.

4.2 Tests of social network predictors' influence on loan default

H_{2a} (social network 1): Loan default is predicted by whether communications are predominantly ingoing or outgoing SMS or voice communications.

Term	Estimate	Std. error	Statistic	<i>p</i> value
(Intercept)	-1.8767	0.0238	-78.8767	0.0000
Incoming Phone	0.3244	0.0255	12.7307	0.0000
Missed phone	0.3742	0.0258	14.5117	0.0000
Outgoing phone	0.4296	0.0245	17.5558	0.0000
SMS inbox	0.2885	0.0239	12.0585	0.0000
SMS sent	0.3082	0.0239	12.8830	0.0000
ACI	2442,161			

H_{2b} (social network 2): Loan default is predicted by social network structure as captured in centrality metrics of the message (either phone or SMS) sender versus receiver.

Term	Estimate	Std. error	Statistic	<i>p</i> value
(Intercept)	-1.3420	0.0026	-516.2669	0.0000
Receiver degree centrality	0.0000	0.0000	-137.2579	0.0000
Sender degree centrality	0.0001	0.0000	77.4380	0.0000
Receiver betweenness centrality	0.0002	0.0000	38.7407	0.0000
Sender betweenness centrality	-0.0005	0.0001	-4.7178	0.0000
AIC	2417,669			

4.3 Tests of geographical landmark proximity and location on the model predictors' influence on loan default

H_{3a} (location): Loan default is predicted by proximity to location-specific business and organizational landmarks. We found the influence of particular classes of landmarks, as evidenced by them being within 50 ms of the point of a call or SMS message location, had a huge impact on default probability and overall AIC improved substantially to 88,607. Table 5 summarizes the findings from our regression of location data.

H_{3b} (loan contract information + location): Loan default is predicted by all of the information in H_1 and H_{3a} . We found that adding loan contract information to the predictors did not substantively change the ranking of landmark proximity from Table 5 on the default rate from that in H_{3b} and overall AIC was slightly worse at 95,209.

H_{3c} (loan contract information + social network + location): Loan default is predicted by all of the information in H_1, H_{2a}, H_{2b} and H_{3a} . We found that adding social network and loan contract information to the predictors did not substantively change the ranking of landmark proximity from Table 5 on the default rate from that in H_{3b} and overall AIC was slightly better at 79,416.

Our regressions under $H_{3a}, H_{3b}, \wedge H_{3c}$ strongly supported the importance of location and proximity to particular sorts of landmarks on the potential for default. Unlike our social networks which were depicted with directed graphs in the study, landmark correlations with higher default do not allow us to identify a causal direction. We arranged Table 5 in decreasing order of likelihood of being associated with a default. Where a borrower had many phone or SMS communications close to a particular landmark, if the GLM coefficient was negative, this would move the default indicator towards '0', i.e., no default indicating a good borrower. Positive coefficients implied the opposite higher risk of default. Lower probability of default (coefficient < -1.5) was correlated with the following landmarks: *city_hall, moving_company, train_station, veterinary_care, parking, natural_feature, car_rental and movie_theater*. Higher probability of default (coefficient > +1.5) was correlated with

the following landmarks: *transit_station*, *park*, *stadium*, *furniture_store*, *library*, *post_office*, *pet_store* and *airport*.

The impact of landmark proximity is highly significant in credit scoring, but explaining why such relationships should exist and speculating ‘why’ these relationships exist would be subjective, and requires additional data in the form of case studies or activity protocols. We are only identifying these landmarks by being within a 50 meter circle of the GPS location (as referenced by Google Maps API) of the phone at the time of the phone call or messaging. From the loan initiators’ perspective, it may not be important to know why the relationship exists, only that it reliably exists.

Table 6 summarizes the fit of each of these nested models using a variety of fit statistics appropriate for logistic regressions. Our preferred model is the one with the lower AIC value. AIC rewards goodness of fit but includes a penalty that is an increasing function of the number of estimated parameters and thus discourages overfitting. AIC measures the quality of a given model in terms of the model’s ability to minimize information loss from the data compared with another model.

Another way to determine the influence of landmarks at particular geographical locations is to analyze the Cook’s distance of each observation [13, 14]. Cook’s distance requires a least squares loss function, and we computed it in conjunction with our synthetic profitability combining information on principal, interest and:

$$\pi = \text{synthetic annual profitability} = \begin{cases} \text{principal} \times \text{interest} & \text{if good loan} \\ 0 & \text{if default} \end{cases}$$

Cook’s distance commonly used to estimate the influence of a data point when performing a least-squares regression analysis and can be interpreted as the distance one’s estimates move within the confidence ellipsoid that represents a region of plausible values for the parameters. In descending sequence, Table 7 shows the predictors that should have the greatest influence on our model’s forecast of profitability. Cooks distance does not differentiate between ‘good’ or ‘bad’ influences; it only addresses the strength of the landmark’s influence on our synthetic profitability. Nonetheless, the most influential of the landmarks shows a striking conformity with the general findings from our ‘default’ effect and logistic regression summarized in Table 5. Thus supports our argument that test results for the simple model using default provide an accurate analysis of the more complex world of practice (see Table 8, and Table 9).

5 Discussions and conclusions

Microlending has grown rapidly and now benefits around 250 million people globally, half who would otherwise not have access to credit. We hoped to provide insight into ways that social credit system can be used to better assess risk and lower the cost while raising availability of lending.

In our dataset of 784 loan contracts, of which 19.39% were in default, representing 21.73% of loaned value, we analyzed a curated dataset of 3577,912 caller

Table 7 Loan default predicted by proximity to location-specific landmarks

Landmark	Est.	p.val	Landmark	Est.	p.val	Landmark	Est.	p.val
(Intercept)	-0.8	0.0	beauty_salon	-0.3	0.0	bowling_alley	0.3	0.6
Campground	-12.4	0.8	Gym	-0.3	0.0	roofing_contractor	0.3	0.8
Synagogue	-11.4	1.0	Finance	-0.3	0.0	Mosque	0.4	0.8
Aquarium	-8.8	1.0	Police	-0.3	0.2	School	0.5	0.0
city_hall	-3.0	0.0	Lawyer	-0.3	0.0	Bakery	0.5	0.0
moving_company	-2.9	0.0	local_government_office	-0.2	0.0	Health	0.5	0.0
train_station	-2.6	0.0	pharmacy	-0.2	0.0	electronics_store	0.5	0.0
veterinary_care	-2.5	0.0	jewelry_store	-0.2	0.0	light_rail_station	0.5	0.0
Parking	-2.4	0.0	Bar	-0.1	0.0	liquor_store	0.5	0.0
natural_feature	-2.3	0.0	Lodging	0.0	0.1	Accounting	0.5	0.0
car_rental	-1.9	0.0	clothing_store	0.0	0.3	shopping_mall	0.6	0.0
movie_theater	-1.7	0.0	meal_takeaway	0.0	0.9	Casino	0.6	0.0
amusement_park	-1.7	0.1	Hospital	0.0	0.9	Laundry	0.6	0.0
Electrician	-1.5	0.2	Florist	0.0	0.9	car_dealer	0.6	0.0
University	-1.4	0.0	Store	0.0	0.2	Museum	0.6	0.0
meal_delivery	-1.4	0.0	Atm	0.0	0.4	Bank	0.6	0.0
funeral_home	-1.4	0.0	insurance_agency	0.0	0.3	Restaurant	0.8	0.0
bus_station	-1.4	0.0	real_estate_agency	0.1	0.0	grocery_or_supermarket	0.8	0.0
taxi_stand	-1.2	0.1	Locksmith	0.1	0.9	convenience_store	1.1	0.0
hardware_store	-1.0	0.0	general_contractor	0.1	0.0	fire_station	1.3	0.0
car_repair	-1.0	0.0	department_store	0.1	0.1	Plumber	1.5	0.2
home_goods_store	-0.9	0.0	travel_agency	0.1	0.0	transit_station	1.6	0.0
Dentist	-0.8	0.0	hair_care	0.1	0.1	Park	1.6	0.0
art_gallery	-0.8	0.0	Neighborhood	0.1	0.0	Stadium	1.8	0.0
night_club	-0.6	0.0	Cafe	0.1	0.0	furniture_store	1.8	0.0
Cemetery	-0.6	0.2	Church	0.1	0.5	Library	1.9	0.0
Food	-0.6	0.0	car_wash	0.1	0.1	post_office	3.3	0.0
Doctor	-0.5	0.0	Storage	0.2	0.7	pet_store	3.8	0.0
Spa	-0.4	0.0	gas_station	0.2	0.0	Airport	4.1	0.0
shoe_store	-0.4	0.0	book_store	0.2	0.0	movie_rental	9.8	0.9
bicycle_store	-0.3	0.2	Embassy	0.2	0.0	Painter	12.1	1.0
			place_of_worship	0.2	0.2	hindu_temple	12.1	0.9

identifiers, aggregated on caller-receiver dyads summing over the call durations to generate an edge-link list where links identify information exchanged between individuals. This is information that in most Western countries would be tightly regulated by government privacy regulations and is not typically available to banks for loan risk assessment. Our access to this information in the Philippines gives us a unique opportunity to gain insights into social credit systems such as China's national system that would not be available with U.S. or European datasets.

Table 8 Fit information criteria for the nested hypothesis tests

	AIC	BIC	deviance	df.residual	null.deviance	df.null	logLik
1a	2300,451	2300,489	2300,445	2654,916	2442,878	2654,918	-1150,223
1b	2248,829	2248,919	2248,815	2654,912	2442,878	2654,918	-1124,408
2a	2442,161	2442,237	2442,149	2654,913	2442,878	2654,918	-1221,074
2b	2417,669	2417,733	2417,659	2654,914	2442,878	2654,918	-1208,829
3a	88,607	89,490	88,417	80,955	105,610	81,049	-44,208
3b	95,209	95,246	95,201	81,046	105,610	81,049	-47,600
3c	79,416	80,328	79,220	80,952	105,610	81,049	-39,610

Table 9 Predictors with the greatest influence on Cook's distance

Predictor	Estimate	Pr(> t)	Predictor	Estimate	Pr(> t)
Mosque	0.025610	0.000000	Florist	0.000104	0.000000
Roofing contractor	0.004506	0.000000	Transit station	0.000079	0.000000
Airport	0.002418	0.000000	Liquor store	0.000059	0.000000
Taxi stand	0.002145	0.000000	Shoe store	0.000057	0.000000
Electrician	0.001384	0.000000	Car rental	0.000044	0.000050
Amusement park	0.001215	0.000000	Parking	0.000031	0.000389
Plumber	0.000853	0.000000	Casino	0.000029	0.000001
Hindu temple	0.000850	0.000000	Shopping mall	0.000025	0.000000
Bowling alley	0.000413	0.000000	Accounting	0.000025	0.000000
Bicycle store	0.000399	0.000000	Hardware store	0.000021	0.000000
Cemetery	0.000393	0.000000	Bar	0.000019	0.000000
Locksmith	0.000372	0.000000	travel agency	0.000017	0.000000
Storage	0.000328	0.000000	University	0.000016	0.000000
Veterinary care	0.000302	0.000000	Department store	0.000016	0.000001
Local government office	0.000175	0.000000	Clothing store	0.000015	0.000000
Museum	0.000162	0.000000	Hair care	0.000014	0.000034
Police	0.000117	0.000000	Home goods store	0.000011	0.000000
Place of worship	0.000111	0.000000	Car repair	0.000008	0.000000
Fire station	0.000109	0.000000			

Exploratory analysis suggested that default was dependent on social network structure as reflected in (1) the 'centrality' of a caller-SMSer; and (2) whether the information was received or sent. Further, we found that good borrowers seemed to communicate more readily with other good borrowers, whereas those who were likely to default tended to communicate with other likely defaulters and social, personal, political and economic similarities are more likely to develop external social networks [1]. Entities are linked together based on common features rather than direct friendship and can be used to predict performance of credit scoring in a smartphone-based microlending setting [28, 52].

Our regressions did not support our initial observations during exploratory analysis concerning the importance of social network structure. Both information on (1) the ‘centrality’ of a caller-messenger; and (2) whether the information was received or sent for our 784 loans provided almost no additional information on potential for default. This information may even have introduced some “noise” into our tests. In a recent P2P lending study, Xu and Chau [72] found that lender-borrower communication significantly influenced funding outcomes, while lender comments reduced the final interest rate by increasing the visibility of a listing and attracting more lenders to bid down the interest rate. This may provide us another perspective to study lender-borrower communications.

Our regressions strongly supported the importance of location and proximity to particular sorts of landmarks on the potential for default. The same person in a different location may act differently, and these actions may be driven by culture. Geographies operating at a transnational scale have profoundly influenced microcredit [11]. Qian et al. [54] found that the impact of geographic proximity on loan price is positively associated with the proximity of borrowers and lenders, and Burtch et al. [9] confirmed evidence that lenders do prefer culturally similar and geographically proximate borrowers. Unlike our social networks which were depicted with directed graphs in the study, landmark correlations with higher default do not allow us to identify a causal direction.

Location is associated with personal preferences [57]. Therefore, geographical targeting can improve the effectiveness of online advertising [42]. Mellinas et al. [45] examined the effects of hotel attribute ratings on location. In online game context, real-life location information could possibly lead people to build up their temporal and emotional involvement toward mobile service [49]. In our study we aim also to look at whether the location is important to predict defaulting loans. In our study, we found that proximity to certain types of landmarks were associated with higher probability of default: *city_hall*, *moving_company*, *train_station*, *veterinary_care*, *parking*, *natural_feature*, *car_rental* and *movie_theater*. Conversely, certain landmarks were correlated with a lower probability of default: *transit_station*, *park*, *stadium*, *furniture_store*, *library*, *post_office*, *pet_store* and *airport*. The differences are significant, but we haven’t tried to elicit any “stories” explaining why such relationships should exist. Indeed, such subjective explanations are beyond the scope of this research and would probably require a significant number of case studies to provide validity. We are only identifying these landmarks by being within a 50-m circle of the GPS location (as referenced by Google Maps API) of the phone at the time of the phone call or messaging. From the loan initiators’ perspective, it may not be important to know why the relationship exists, only that it reliably exists.

6 Contributions and limitations

There are several practical contributions that should be mentioned. Mobile social credit platforms allow easy installation of automated credit scoring for evaluating purposes, which in turn may reduce cost of credit analysis, improve cash flow, enable faster credit decisions, and reduce losses [36]. Through our findings, microfinance

firms can build a better automatic credit scoring system by incorporating location and loan contract information. Managers of social credit platforms can also see benefits by carrying out different factors which better develop credit scoring for discriminating between good and bad loans. Information asymmetries are a significant challenge in emerging economies; borrowers know their own financial situation, while lenders know very little about borrowers [72]. To overcome these challenges, firms can substitute for loan histories, data from social webs (social activity, social capital), demography (location, age, gender, marital status) and consumer communication (SMS and voice communications) data to help better predict lending performance.

Our study has several implications for credit scoring practice and research. First, in a financial or accounting context, the focus is typically on earnings of default models when considering credit, and the models typically are credit scoring models with particular predictors related to prior behavior, finances and lifestyle. We extended those predictors to include communications on mobile platforms. There are many modes of communications, with an evolving landscape that increasingly favors visual communications. Microlending has been sporadically studied and much of the existing literature addresses corporate and bank credit. Our research adds to this line of study, and makes contributions that have not been previously shown. Secondly, datasets such as those that were provided are typically held by the managers of the database systems with privacy concerns being paramount. The two streams of communication used for this purpose are textual, and probably the major text communications channels used on mobile platforms are for the particular geographical locations of the system providers. There may be important additional information in pictures and snapshots on other platforms, but these are difficult to interpret. Given the dearth of studies in microlending credit scoring, we felt that the rather extensive work we have communicated in this paper marks a substantial contribution to the existing literature.

Acknowledgements We would like to thank the editor and anonymous reviewers for their comments, which have greatly improved our paper. This study supported by the Fundamental Research Funds for the Central Universities of No. (BX180604).

Compliance with ethical standards

Conflict of interest There is no conflict of interest for this study.

References

1. Adler, P. S., & Concept, N. (2002). Social capital: Prospects for a new concept. *The Academy of Management Review*, 27(1), 17–40.
2. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Paper presented at the 2nd international symposium on information theory, Akademiai Kiado, Budapest, 1973*.
3. Akaike, H. (1998). A Bayesian analysis of the minimum AIC procedure. *Selected papers of Hirotugu Akaike* (pp. 275–280). Springer.

4. Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2012). Analysis of ego network structure in online social networks. *Paper presented at the privacy, security, risk and trust (PASSAT), 2012 international conference on and 2012 international conference on social computing (SocialCom)*.
5. Barabási, A. L. (2005). The origin of bursts and heavy tails in human dynamics. arXiv preprint cond-mat/0505371.
6. Barabási, A. L. (2007). *Network medicine—from obesity to the “diseasome”*. Waltham: Mass Medical Soc.
7. Böhme, R., & Pötzsch, S. (2010). Privacy in online social lending. *Paper presented at the AAAI spring symposium: Intelligent information privacy management*.
8. Borgatti, S. P., Mehra, A., Brass, D. J., & Lopianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892–895.
9. Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural differences and geography as determinants of online prosocial lending. *MIS Quarterly*, 38(3), 773–794.
10. Callaway, D. S., Newman, M. E., Strogatz, S. H., & Watts, D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25), 5468.
11. Carr, J., Dickinson, E., McKinnon, S. L., & Chávez, K. R. (2016). Kiva's flat, flat world: Ten years of microcredit in cyberspace. *Globalizations*, 13(2), 143–157.
12. Cohen, R., & Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5), 058701.
13. Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.
14. Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74(365), 169–174.
15. Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
16. Csárdi, G., & Nepusz, T. (2010). igraph reference manual. <http://igraph.sourceforge.net/documentation.html>. Accessed 20 April.
17. Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2018). The microfinance business model: Enduring subsidy and modest profit. *The World Bank Economic Review*, 32(2), 221–244.
18. Daft, R. L., & Robert, H. L. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554–571.
19. de Nooy, W. (2012). Graph theoretical approaches to social network analysis. In *Computational complexity: Theory, techniques, and applications* (pp. 2864–2877). Heidelberg: Springer.
20. Dennis, A. R., Robert, M. F., & Joseph, S. V. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, 32(3), 575–600.
21. de Soto, H. (2014). Missing ingredients of globalization. In *The future of globalization* (pp. 37–51). Abingdon: Routledge.
22. de Soto, H. (2017). A tale of two civilizations in the era of Facebook and blockchain. *Small Business Economics*, 49(4), 729–739.
23. Dillon, T. W., & Lending, D. (2010). Will they adopt? Effects of privacy and accuracy. *Journal of Computer Information Systems*, 50(4), 20–29.
24. Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: Cambridge University Press.
25. Ebel, H., Mielsch, L. I., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, 66(3), 035103.
26. Everett, M., & Borgatti, S. P. (2005). Ego network betweenness. *Social Networks*, 27(1), 31–38.
27. Fernandes, G. B., & Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, 249(2), 517–524.
28. Godlewski, C. J., Sanditov, B., & Burger-Helmchen, T. (2012). Bank lending networks, experience, reputation, and borrowing costs: Empirical evidence from the French syndicated lending market. *Journal of Business Finance & Accounting*, 39(1), 113–140.
29. Grodzinsky, F. S., & Tavani, H. T. (2005). P2P networks and the Verizon v. RIAA case: Implications for personal privacy and intellectual property. *Ethics and Information Technology*, 7(4), 243–250.
30. Hoogeveen, J. G. M. (2002). Income risk, consumption security and the poor. *Oxford Development Studies*, 30(1), 105–121.

31. Huang, Y. L. (2009). Prediction of contractor default probability using structural models of credit risk: An empirical investigation. *Construction Management and Economics*, 27(6), 581–596.
32. Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale JL & Tech.*, 18, 148.
33. Jalali, M. S., Ashouri, A., Herrera-Restrepo, O., & Zhang, H. (2016). Information diffusion through social networks: The case of an online petition. *Expert Systems with Applications*, 44, 187–197.
34. Jones, C., & Volpe, E. H. (2011). Organizational identification: Extending our understanding of social identities through social networks. *Journal of Organizational Behavior*, 32(3), 413–434.
35. Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. Oxford: OUP USA.
36. Kiruthika, & Dilsha, M. (2015). A neural network approach for microfinance credit scoring. *Journal of Statistics and Management Systems*, 18(1–2), 121–138.
37. Lacker, J. M. (2002). The economics of financial privacy: To opt out or opt in? *Economic Quarterly-Federal Reserve Bank of Richmond*, 88(3), 1–16.
38. Lawrence, E. C., Smith, L. D., & Rhoades, M. (1992). An analysis of default risk in mobile home credit. *Journal of Banking & Finance*, 16, 299–312.
39. Lee, C. H., & Chiravuri, A. (2019). Dealing with initial success versus failure in crowdfunding market: Serial crowdfunding, changing strategies, and funding performance. *Internet Research*. <https://doi.org/10.1108/INTR-03-2018-0132>.
40. Leskovec, J., & McAuley, J. J. (2012). Learning to discover social circles in ego networks. *Paper presented at the Advances in neural information processing systems*.
41. Leskovec, J., & Sosić, R. (2016). Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1), 1.
42. Lian, S., Cha, T., & Xu, Y. (2019). Enhancing geotargeting with temporal targeting, behavioral targeting and promotion for comprehensive contextual targeting. *Decision Support Systems*, 117, 28–37.
43. Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Paper presented at the Carnegie-Rochester conference series on public policy*.
44. McCord, G. C., & Sachs, J. D. (2015). Physical geography and the history of economic development.
45. Mellinas, J. P., Nicolau, J. L., & Park, S. (2019). Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. *Tourism Management*, 71, 421–427.
46. Meissner, M. (2017). China's social credit system: A big-data enabled approach to market regulation with broad implications for doing business in China. *Mercator Institute for China studies*, 24, 1–13.
47. Mimouni, K. (2017). Currency risk and microcredit interest rates. *Emerging Markets Review*, 31, 80–95.
48. Morduch, J., Cull, R., & Demirgüç-Kunt, A. (2017). The microfinance business model: Modest profit and enduring subsidy. *World Bank Economic Review*.
49. Oh, Y. J., Park, H. S., & Min, Y. (2019). Understanding location-based service application connectedness: Model development and cross-validation. *Computers in Human Behavior*, 94, 82–91.
50. Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., et al. (2007). Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6), 179.
51. Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., et al. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18), 7332–7336.
52. Óskarsdóttir, M., Bravo, C., Sarraute, C., Baesens, B., & Vanthienen, J. (2018). Credit scoring for good: Enhancing financial inclusion with smartphone-based microlending. In *the 39th international conference on information systems*, San Francisco.
53. Price, D. J. D. S. (1965). Networks of scientific papers. *Science*, 149(3683), 510–515.
54. Qian, X., Kong, D., & Du, L. (2019). Proximity, information, and loan pricing in internal capital markets: Evidence from China. *China Economic Review*, 54, 434–456.
55. Riggins, F. J., & Weber, D. M. (2017). Information asymmetries and identification bias in P2P social microlending. *Information Technology for Development*, 23(1), 107–126.
56. Sachs, J. D. (2015). *The age of sustainable development*. New York: Columbia University Press.
57. Samoggia, A., & Riedel, B. (2018). Coffee consumption and purchasing behavior review: Insights for further research. *Appetite*, 129, 70–81.

58. San Pedro, J., Proserpio, D., & Oliver, N. (2015). MobiScore: Towards universal credit scoring from mobile phone data. *Paper presented at the international conference on user modeling, adaptation, and personalization*.
59. Sanchez, P., Palm, C., Sachs, J., Denning, G., Flor, R., Harawa, R., et al. (2007). The African millennium villages. *Proceedings of the National Academy of Sciences*, 104(43), 16775–16780.
60. Scott, J. (2017). *Social network analysis*. Thousand Oaks: Sage.
61. Scott, W. R., & Davis, G. F. (2003). Networks in and around organizations. *Organizations and Organizing*. Pearson Prentice Hall.
62. Serrano-Cinca, C., Gutiérrez-Nieto, B., & Reyes, N. M. (2016). A social and environmental approach to microfinance credit scoring. *Journal of Cleaner Production*, 112, 3504–3513.
63. Shi, W. (2015). Internet lending in China: Status quo, potential risks and regulatory options. *Computer Law & Security Review*, 31, 793–809.
64. Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825), 268.
65. Tang, S., & Guo, S. (2017). Formal and informal credit markets and rural credit demand in China. *Paper presented at the industrial economics system and industrial security engineering (IEIS'2017), 2017 4th international conference on*.
66. Tsarenko, Y., & Roosrani Tojib, D. (2009). Examining customer privacy concerns in dealings with financial institutions. *Journal of Consumer Marketing*, 26(7), 468–476.
67. Uysal, V. B., Kedia, S., & Panchapagesan, V. (2008). Geography and acquirer returns. *Journal of Financial Intermediation*, 17, 256–275.
68. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.
69. Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit scoring with social network data. *Marketing Science*, 35(2), 234–258.
70. Wikipedia, Maximum likelihood estimation, https://en.wikipedia.org/wiki/Maximum_likelihood_estimation.
71. Xia, Y., Chi, K. T., Tam, W. M., Lau, F. C., & Small, M. (2005). Scale-free user-network approach to telephone network traffic analysis. *Physical Review E*, 72(2), 026116.
72. Xu, J. J., & Chau, M. (2018). Cheap talk? The impact of lender borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems*, 35(1), 53–85.
73. Yan, J., Wang, K., Liu, Y., Xu, K., Kang, L., Chen, X., et al. (2018). Mining social lending motivations for loan project recommendations. *Expert Systems with Applications*, 111, 100–106.
74. Yunus, M. (1999). The Grameen bank. *Scientific American*, 281(5), 114–119.
75. Yunus, M. (2007). *Banker to the poor*. New Delhi: Penguin Books India.
76. Yunus, M. (2009). *Creating a world without poverty: Social business and the future of capitalism*. New York: Public Affairs.
77. Zhang, K., & Zhang, F. (2016). Report on the construction of the social credit system in China's Special Economic Zones. *Annual report on the development of China's Special Economic Zones (2016)* (pp. 153–171). Springer.
78. Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2016). Research on credit scoring by fusing social media information in online peer-to-peer lending. *Procedia Computer Science*, 91, 168–174.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Jian Mou¹  · J. Christopher Westland² · Tuan Q. Phan³ · Tianhui Tan³

J. Christopher Westland
westland@uic.edu

Tuan Q. Phan
tuan.quang.phan@gmail.com

Tianhui Tan
tinweitam@gmail.com

- ¹ School of Business, Pusan National University, Busan, Republic of Korea
- ² University of Illinois – Chicago, Chicago, USA
- ³ National University of Singapore, Singapore, Singapore