

Chen, Xiao; Chong, Zhaohui; Giudici, Paolo; Huang, Bihong

Working Paper

Networking with peers: Evidence from a P2P lending platform

ADB Working Paper Series, No. 1080

Provided in Cooperation with:

Asian Development Bank Institute (ADBI), Tokyo

Suggested Citation: Chen, Xiao; Chong, Zhaohui; Giudici, Paolo; Huang, Bihong (2020) : Networking with peers: Evidence from a P2P lending platform, ADB Working Paper Series, No. 1080, Asian Development Bank Institute (ADBI), Tokyo

This Version is available at:

<https://hdl.handle.net/10419/238437>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/>



ADB Working Paper Series

**NETWORKING WITH PEERS:
EVIDENCE FROM A P2P
LENDING PLATFORM**

Xiao Chen, Zhaohui Chong,
Paolo Giudici, and Bihong Huang

No. 1080
February 2020

Asian Development Bank Institute

Xiao Chen is a postdoctoral fellow at the Chinese University of Hong Kong. Zhaohui Chong is an assistant professor at Shangtou University. Paolo Giudici is a professor of statistics at the University of Pavia. Bihong Huang is a research fellow at the Asian Development Bank Institute.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

The Asian Development Bank refers to "China" as the People's Republic of China.

Suggested citation:

Chen, X., Z. Chong, P. Giudici, and B. Huang. 2020. Networking with Peers: Evidence from a P2P Lending Platform. ADBI Working Paper 1080. Tokyo: Asian Development Bank Institute. Available: <https://www.adb.org/publications/networking-peers-evidence-p2p-lending-platform>

Please contact the authors for information about this paper.

Email: Chenxiao427@163.com, zhchong@stu.edu.cn, paolo.giudici@unipv.it,
bihuang@adbi.org

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2020 Asian Development Bank Institute

Abstract

The paper investigates the role of network centrality in predicting borrowers' and lenders' behavior in peer-to-peer (P2P) lending. The empirical analysis of data from Renrendai, a leading lending platform in the People's Republic of China, reveals that the lenders who are at the center of a network not only invest larger amounts but also invest more swiftly than their peers, reflecting the information advantage arising from their position in the network. Furthermore, the borrowers who are at the center of a network are able to borrow at lower interest rates and with higher success rates. At the same time, they are less likely to default. These findings imply that, in the P2P lending market, network linkages not only enhance market efficiency but also encourage reputation protection.

Keywords: network centrality, social networks, peer-to-peer lending, lending behavior, loan outcomes

JEL Classification: C21, C58, E44, G21

Contents

1.	INTRODUCTION	1
2.	LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT	4
2.1	Literature Review	4
2.2	Hypothesis Development	6
3.	DATA AND METHODOLOGY	7
3.1	Data	7
3.2	Measurement of Key Variables	8
3.3	Summary Statistics	11
3.4	Network-Based Regression	15
4.	EMPIRICAL ANALYSIS	16
4.1	Evolution of the P2P Network	17
4.2	Network and Lending Behavior	19
4.3	Network and Borrowing Outcome	22
4.4	Addressing the Endogeneity Concerns	28
5.	CONCLUSIONS	41
	REFERENCES	42

1. INTRODUCTION

Research has increasingly recognized the importance of social networks in traditional financial markets due to their role in shaping the judgment and decision making of firms (Khanna, Kim, and Lu 2015; Bajo et al. 2016), banks (Iyer and Puri 2012; Grullon, Underwood, and Weston 2014), venture capital (VC) (Hochberg, Ljungqvist, and Lu 2007, 2010), individual investors (Ozsoylev et al. 2014; Hong and Xu 2019), mutual fund investment (Hong, Kubik, and Stein 2005; Cohen, Frazzini, and Malloy 2008), and financial analysts (Fang and Huang 2017). Online peer-to-peer (P2P) lending has recently emerged as an appealing alternative to traditional bank lending around the world (Sorenson et al. 2016; Wei and Lin 2017). Despite the intensive interactions among the participants on the platforms, our knowledge on the position of different borrowers and lenders in P2P lending networks and its potential impact on the behavior of the participants is still limited. This paper attempts to fill this gap by applying social network analysis (SNA) to gauge the centrality of borrowers and lenders in P2P lending and to investigate its influence on the market outcome.

Compared with traditional bank lending, P2P lending is easy to access, enhances financial inclusion, and improves funding efficiency (Iyer et al. 2016; Paravisini, Rappoport, and Ravina 2017). Borrowers can post loan requests without providing collateral or guarantees, while lenders can make lending decisions according to the information that the borrowers disclose. Bypassing banks, both lenders and borrowers are anonymous and do not establish physical connections. This may help to alleviate the potential concerns of relationship lending that are common in traditional credit markets. However, a new type of social network has arisen through the interaction among the participants in P2P lending, and its effects on the market outcome remain underexplored.

A growing amount of literature has documented the role of social networks, particularly the role of information regarding friends and groups in the P2P lending market, in moderating the information asymmetry between borrowers and lenders (Pope and Sydnor 2011; Lin, Prabhala, and Viswanathan 2013; Liu et al. 2015; Iyer et al. 2016; Hildebrand, Puri, and Rocholld 2017). For example, Lin, Prabhala, and Viswanathan (2013) found that online friendship among borrowers acts as a signal of credit quality. Liu et al. (2015) showed that friendship among borrowers, especially among close offline friends, facilitates lending. The empirical evidence that Lin and Viswanathan (2016) provided suggests that home bias is a robust phenomenon, even in the context of a large online crowdfunding marketplace. Our study extends the existing research by applying social network analysis to quantify the centrality of borrowers and lenders and identify its impact on market outcomes. The network centrality of lenders influences their investment decision by increasing the flow of information and encouraging risk taking. A lender who is centrally located in a network is able to receive information signals earlier than his or her peers and is more likely to obtain valuable information to infer the creditworthiness of borrowers. Such an information advantage enables lenders with a higher level of network centrality not only to invest more but also to make investment decisions faster than others. At the same time, the network centrality helps borrowers to improve their loan requesting strategy, enhance their incentive for reputation protection, and moderate moral hazard. A higher level of network centrality makes borrowers more knowledgeable about the market conditions and allows them to set loan requests that are more consistent with lenders' needs, enhancing the probability of funding success.

Using data from Renrendai, a leading Chinese P2P lending platform, we map the interactions among the platform participants by gauging the position of each lender and borrower in the network. We construct five centrality measures that the social network

analysis literature has widely used: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*. Accordingly, we consider both direct and indirect connections. We individuate an indirect social network connection between two lenders when they invest in the same borrower. If a lender invests in multiple borrowers, he or she will extend his network to multiple lenders. The direct connections between lenders and the indirect connection between lenders through borrowers form the lender's social network. Similarly, the direct connections between borrowers and the indirect connections between borrowers through lenders form the borrowers' social network. Our empirical results show that, while the number of participants in the platform has increased, the number of connections has not grown correspondingly. This means that a smaller number of lenders have co-invested in loan listings and that their investment portfolios have a high degree of concentration.

We investigate the impact of network centrality on lenders' investment behavior. The empirical analysis of our unique database consisting of 273,508 bidding records on 116,460 loan listings generates several new and interesting findings. We find that a lender who is at the center of a P2P network not only invests a larger amount but also bids more quickly for a loan than his or her peers. Our finding echoes Faley, Kovacs, and Venkateswaran's (2014) claim that personal social networks allow lenders to access various sources of information. This strengthens their confidence when they make investment decisions, as they are willing to take more risks. Moreover, lenders who are at the center of a network usually have richer investment experiences, which enable them to act as leading lenders, rather than followers, in the bidding process.

We also examine the impact of network centrality on borrowers' behavior. Our empirical results show that borrowers with higher centrality are more likely to borrow at lower interest rates. This finding is consistent with the existing research, acknowledging that borrowers with higher centrality have greater bargaining power and better market insight and therefore are more sophisticated in setting interest rates (Hochberg, Ljungqvist, and Lu 2010). Moreover, we find that borrowers with higher centrality tend to have higher funding success rates but lower default rates. Similar to central lenders, central borrowers accumulate richer knowledge and experience in raising funds through P2P lending platforms. It is easier for them to set loan requests consistent with lenders' expectations and requirements. In addition, the negative relationship between centrality and default reflects that the borrowers at the center of the network care more about their reputation and have lower moral hazard than their peers. This may be due, on one hand, to the fact that borrowers with higher network centrality have accumulated more social capital in the P2P lending market, which makes them more attractive to lenders; on the other hand, a default would seriously destroy a borrower's social capital. Rational borrowers protect their reputation by avoiding default.

We believe that our paper contributes to the burgeoning research in the field of financial technology, and it extends different strands of related literature. First, the paper contributes to the growing stream of research applying social network analysis to different areas of finance, including venture capital (VC) (Hochberg, Ljungqvist, and Lu 2007), mergers and acquisitions (Cai and Sevilir 2012; El-Khatib, Fogel, and Jandik 2015), mutual fund investment (Hong, Kubik, and Stein 2005; Cohen, Frazzini, and Malloy 2008), CEO compensation (Butler and Gurun 2012; Engelberg, Gao, and Parsons 2013; Shue 2013), initial public offerings (IPOs) (Grullon, Underwood, and Weston 2014; Bajo et al. 2016), corporate governance (Fracassi and Tate 2012; Khanna, Kim, and Lu 2015), innovation (Faley, Kovacs, and Venkateswaran 2014), individual investors (Ozsoylev et al. 2014; Hong and Xu 2019), and bank lending (Garmaise and Moskowitz 2003; Houston, Lee, and Suntheim 2018). Ozsoylev et al. (2014) confirmed that information diffusion among investors influences trading behavior and returns. Houston,

Lee, and Suntheim (2018) showed that banks with shared social connections partner more often in the global syndicated loan market and that central banks in the network play dominant roles in various interbank transactions, signifying the role of social connections in facilitating business connections. Despite the intensive interactions among the participants in P2P lending, the knowledge on the position of borrowers and lenders on the platform and its impacts on the market equilibrium is scarce. We fill this gap by extending social network analysis to P2P lending, which is a rapidly growing alternative to traditional lending institutions.

Second, we enrich the literature investigating the role of social networks in facilitating financial transactions (Stuart and Yim 2010; Nguyen 2012; Fracassi 2017; Hong and Xu 2019). Hochberg, Ljungqvist, and Lu (2007) suggested that better-networked VC firms experience significantly better fund performance, which they measured using the proportion of investments that firms successfully dispose of through an IPO or a sale to another company, reflecting that syndicated networks facilitate the sharing of information, contacts, and resources among VCs. Cohen, Frazzini, and Malloy (2010) demonstrated the impact of social networks on agents' ability to gather superior information about firms. Duchin and Sosyura (2013) found that divisional managers with social connections to the CEO receive more capital, because the social connections improve the quality of information about the division's investment opportunities. Pool, Stoffman, and Yonker (2015) showed that socially connected fund managers have more similar holdings and trades. This is in line with the reality that managers who live closer to each other have a better chance of meeting and subsequently becoming acquaintances or friends, which enables the flow of information through informal person-to-person relationships. These studies have emphasized social networks as a clue to the dissemination of knowledge, ideas, or private information, which has important implications for financial transactions. Our study underlines the importance of the information channel through which network centrality affects a lender's (or a borrower's) behavior in P2P lending.

Third, we contribute to the booming peer-to-peer lending literature (Fuster et al. 2019; Tang 2019; Vallée and Zeng 2019) by introducing a new perspective on social network analysis. Despite the explosive growth of peer-to-peer lending, information asymmetry remains a critical issue, and this evolving credit market is likely to amplify it compared with a traditional credit market. Existing research has investigated a wide range of mechanisms to assess the creditworthiness of borrowers in P2P lending markets. For example, using data from Prosper.com, the leading P2P lending platform in the US, Duarte, Siegel, and Young (2012) showed that borrowers who appear to be more trustworthy are more likely to receive funding following their borrowing requests. Michels (2012) confirmed that borrowers who disclose more information voluntarily are more likely to borrow at lower rates. Some research has shown that a borrower's social network reflects his or her creditworthiness (Lin, Prabhala, and Viswanathan 2013). To the best of our knowledge, this paper is the first attempt to apply the social network analysis approach to measure the network centrality of P2P lending participants and investigate its impact on market outcomes. We confirm that lenders who are better connected are more confident in their investment decisions and are willing to take more risks, while borrowers with a higher degree of centrality gain stronger bargaining power and care more about their reputation. Our study sheds new light on the behavior of P2P participants in their own networks.

We organize the rest of this paper as follows: section 2 reviews the relevant literature and develops our research hypothesis; section 3 describes our measurement model; section 4 details the main empirical analysis and related robustness exercises; and section 5 concludes the paper with a discussion.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

In this section, we review the relevant literature and derive a verifiable hypothesis on the role of social networks in P2P lending.

2.1 Literature Review

The employment of centrality analysis to study the behavior and influence of individuals in a network has been gaining increasing momentum in economics and finance over the last decade (Borgatti et al. 2009; El-Khatib, Fogel, and Jandik 2015; Bourles, Bramouille, and Perez-Richet 2017; Cruz, Labonne, and Querubin 2017). For example, El-Khatib, Fogel, and Jandik (2015) found that network centrality allows CEOs to gather and control private information efficiently, facilitating value-creating acquisition decisions. Using various centrality measures from SNA, Bajo et al. (2016) investigated how the location of a lead IPO underwriter in a network of investment banks affects various IPO characteristics. Cruz, Labonne, and Querubin (2017) claimed that candidates for public offices are disproportionately drawn from more central families and that family network centrality contributes to larger vote shares during the elections. Following a similar approach, Rossi et al. (2018) inferred a positive relationship between network centrality and risk-adjusted performance in a delegated investment management setting.

Network centrality can be measured in a variety of ways to capture different dimensions of the network. The most-used measures of network centrality are degree, betweenness, eigenvectors, and closeness. Bajo et al. (2016) essentially measured the number of ties of an underwriter with other investment banks. El-Khatib, Fogel, and Jandik (2015) used BoardEx data to construct a social network of CEOs of American firms and calculate degree, betweenness, eigenvector, and closeness centrality measures for all the individuals connected in the network. Cruz, Labonne, and Querubin (2017) applied eigenvector centrality to capture a political candidate's family network in municipal elections. Houston, Lee, and Suntheim (2018) employed degree, betweenness, eigenvector, and closeness measures to reflect the connections among the board members of different banks. Following the literature in this strand, we construct five centrality measures in the context of a P2P lending network: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*.

Social networks play an important role in shaping the behavior of market participants. In recent years, with the advancement of measurement methods and the increasing availability of financial micro-data, many empirical studies have emerged to examine the impact of social networks on corporate governance, M&As, venture capital, corporate investment, individual investors' behavior, bank credit, and other micro-financial activities (Cohen, Frazzini, and Malloy 2008; Han and Yang 2013; Khanna, Kim, and Lu 2015; Bajo et al. 2016; Ahern 2017; Cruz, Labonne, and Querubin 2017; Fracassi 2017; Houston, Lee, and Suntheim 2018). Visualizing the relationship between the members of a social network as a structure of nodes and ties (Freeman 1978; El-Khatib, Fogel, and Jandik 2015; Rossi et al. 2018), these studies found that social networks spur information flows across individuals and organizations, provide channels and bridges for individuals to interact with each other (Granovetter 1973), exchange information and resources, reinforce existing relationships, establish new relationships, and, therefore, reshape the economic behavior of different members of the network. Other papers in the same stream include Cohen, Frazzini, and Malloy (2010); Shue (2013); Pool, Stoffman, and Yonker (2015); and Calomiris and Carlson (2017), among others.

In bank lending, social networks affect a borrower's access to finance and the borrowing costs. Garmaise and Moskowitz (2003) found that service intermediaries, although they do not supply loans by themselves, can facilitate their clients' access to finance through informal relationships with the lenders. This result implies that, even in the United States, where capital markets are well developed, access to finance still largely depends on informal networks. Engelberg, Gao, and Parsons (2012) showed that interest rates decline remarkably when interpersonal linkages, such as university or college alumni relations, connect the managers of banks and firms. Using a large sample of syndicated loans, Ferreira and Matos (2012) investigated the effects of bank controls on borrower firms through their representation on the boards of directors or by the holding of shares through bank asset management divisions and claimed that these linkages mitigate credit rationing effects during crisis times. Haselmann, Schoenherr, and Vig (2018) examined the credit allocation decisions of banks to firms inside a network and concluded that elite networks act as rent extractive coalitions that stifle economic prosperity. Houston, Lee, and Suntheim (2018) showed that banks with shared social connections partner more often in the global syndicated loan market. In addition, the banks in the central part of a network play dominant roles in various interbank transactions, indicating that social connections facilitate business connections. Stanton, Walden, and Wallace (2018) asserted that default risk, which is closely related to network positions, evolves predictably among linked nodes. Furthermore, the loan quality estimated from a network model is correlated with independent quality assessments, altogether pointing to the vital importance of network effects in the US mortgage market.

In P2P lending, an alternative to bank lending, interactions among participants have exhibited a powerful influence on loan outcomes. Pope and Sydnor (2011) concluded that group membership exerts extra social pressure on loan repayment when the repayment activities of its members affect the group rating. Based on data from Prosper.com, a leading online lending platform in the United States, Lin, Prabhala, and Viswanathan (2013) found that the online friendships of borrowers, which show a striking gradation based on the roles and identities of the friends, act as signals of credit quality by increasing the funding probability, lowering borrowing rates, and reducing ex post default rates. Employing transaction data from PPDai.com, one of the major online lending platforms in the People's Republic of China (PRC), Liu et al. (2015) investigated the different roles of friendship, including pipes, prisms, and herding signals, in shaping loan outcomes. Their empirical results showed that, although the friends of a borrower, especially close offline friends, act as financial pipes by lending money to the borrower, their endorsements via online bidding on a loan negatively affect subsequent bids by third parties, reflecting the prism effect of friendship. However, a potential lender tends to follow his offline friends, especially close friends, to place a bid, reflecting a relational herding effect among lenders.

However, few studies have examined the role of a participant's network centrality in P2P lending. In particular, research on lenders' social networks is very scarce. This paper attempts to shed light on the role of network centrality in P2P lending from both lenders' and borrowers' perspectives. Although many borrowers (lenders) appear to have independent relationships with other borrowers (lenders) at the individual level, they are connected directly or indirectly through networks. In theory, a central position in the network enables market participants to gain more advantages in accessing information, disseminating knowledge, establishing a reputation, or accumulating social capital (Ozsoylev et al. 2014; Pool, Stoffman, and Yonker 2015; Kilduff et al. 2016; Ahern 2017; Breza and Chandrasekhar 2019). Based on social network analysis, we first assess the impact of lenders' network position on their investment decisions. We then examine the

role of borrowers' network position in shaping their loan requesting strategy and loan outcomes.

2.2 Hypothesis Development

The existing literature has clearly acknowledged the role of social networks in facilitating information exchange and spurring the transmission of knowledge and ideas (Hong, Kubik, and Stein 2005; Brown et al. 2008; Cohen, Frazzini, and Malloy 2008; Moretti 2011; Duchin and Sosyura 2013; El-Khatib, Fogel, and Jandik 2015; Pool, Stoffman, and Yonker 2015; Cai and Szeidl 2018; Stanton, Walden, and Wallace 2018). Cohen, Frazzini, and Malloy (2008) recognized social networks as an important mechanism for information to flow into asset prices. Engelberg, Gao, and Parsons (2012) presented evidence that social networks lead to either a better information flow or better monitoring. El-Khatib, Fogel, and Jandik (2015) measured the extent and strength of CEOs' personal connections and concluded that high network centrality allows CEOs to gather and control private information efficiently, facilitating value-creating acquisition decisions. Assuming that the investment banking network plays the role of information dissemination and information extraction, Bajo et al. (2016) investigated how the location of a lead initial public offering (IPO) underwriter in its network of investment banks affects various IPO characteristics. Ahern (2017) empirically showed that inside information flowing through strong social ties based on family, friends, and geographic proximity enables inside traders to earn prodigious returns of 35% over 21 days, with more central traders earning greater returns, as information conveyed through social networks improves price efficiency.

Without financial intermediation, the issue of information asymmetry is serious in P2P lending markets. Lenders have to rely heavily on the information that borrowers reveal to infer their creditworthiness. Exploring the various mechanisms that could moderate the information asymmetry between borrowers and lenders is critical (Strausz 2017). In P2P lending, social networks facilitate the sharing of information, contacts, and resources among the participants. Lenders who are centrally located in such a network tend to receive more information to infer borrowers' creditworthiness, which enables them to make investment decisions more swiftly.

Following the previous literature, we put forward the following hypothesis:

Hypothesis 1: *The network centrality of the lenders on a P2P lending platform significantly affects their investment behavior through the information channel.*

Personal social networks also affect individual behavior by changing the level of risk taking, reputation protection, and moral hazard (Kuhnen 2009; Fracassi and Tate 2012; Feigenberg, Field, and Pande 2013; Ambrus, Mobius, and Szeidl 2014; Khanna, Kim, and Lu 2015; Schmidt 2015; Munshi and Rosenzweig 2016; Bourles, Bramoulle, and Perez-Richet 2017; Breza and Chandrasekhar 2019; Schoenherr 2019). Karlan et al. (2009) asserted that network connections between individuals can act as social collateral to secure informal borrowing. Fracassi and Tate (2012) pointed out that network ties with the CEO weaken the intensity of board monitoring. Faley, Kovacs, and Venkateswaran (2014) postulated that two characteristics of personal networks alleviate CEO risk aversion in investment decisions. On one hand, personal connections increase CEOs' access to relevant network information, which encourages innovation by helping to identify, evaluate, and exploit innovative ideas; on the other hand, personal connections provide CEOs with labor market insurance that facilitates investments in risky innovations by mitigating the career concerns inherent in such investments. Ishii and Xuan (2014) confirmed that social ties between the acquirer and the target may lead to poorer decision making and lower value creation for

shareholders. Khanna, Kim, and Lu (2015) argued that the connections that CEOs develop with top executives and directors through their appointment decisions increase the risk of corporate fraud. Hasan et al. (2017) noted that debt holders perceive social capital as providing environmental pressure that constrains opportunistic firm behaviors in debt contracting. Gurun, Stoffman, and Yonker (2018) provided competing evidence that social networks can transmit trust shock. Rossi et al. (2018) showed a positive relationship between network centrality and risk-adjusted performance in a delegated investment management setting. More connected managers take more portfolio risk and receive larger investor flows because they are able to exploit investment opportunities through their network connections.

Similar to central lenders, borrowers with a higher degree of network centrality accumulate richer knowledge and experience in raising funds through P2P lending platforms. We argue that higher network centrality improves borrowers' market insight, making them more sophisticated in setting loan requests. Moreover, borrowers with higher network centrality accumulate more social capital in the P2P lending market, which makes them more attractive to lenders. This also implies that a default would seriously destroy a borrower's social capital. Rational borrowers should protect their reputations by avoiding default. Following the previous literature, we put forward the following hypothesis:

Hypothesis 2: *The network centrality of the borrowers on a P2P lending platform significantly affects the loan outcome and loan performance.*

Both hypotheses, if verified, will advance the existing literature, leading to further and more precise insights into network effects on peer-to-peer borrowing and lending. In the next section, we present the data and the statistical methodology that we employ in the empirical analysis aimed at verifying our two research hypotheses.

3. DATA AND METHODOLOGY

3.1 Data

We obtained the data used in this study from Renrendai, one of the largest peer-to-peer lending platforms in the PRC. Founded in 2010, it currently has over 1 million members located in more than 2,000 cities, towns, and counties across the country. Renrendai has received recognition for its reputation: in 2014 and 2015, the Internet Society of China and the China Academy of Social Science awarded it a rating of AAA (the highest level) as an online lending platform. In 2015, it ranked no. 53 on a list of the PRC's top 100 internet companies that the Internet Society of China and the Ministry of Industry and Information released.

The transactions taking place at Renrendai are typical P2P lending transactions. Borrowers post loan listings with information about their borrowing purpose, amount, interest rate, and term. Renrendai only provides basic verification of borrowers' national identification cards, credit reports, and locations. Akin to Prosper.com, Renrendai's profit mainly comes from borrowers' closing fees and lenders' servicing fees. Once a customer posts a loan listing online, lenders may place bids by stating the amount that they want to fund. With a minimum bid amount of RMB 50, a listing typically requires dozens of bids to become fully funded. A listing that achieves 100% funding is a successful listing; otherwise, the borrower receives no funding.

We collected all the loan listings appearing on Renrendai between 1 January 2011 and 31 December 2015. For each loan listing, we have the borrower's ID, borrowing amount,

interest rate, term, and corresponding bidding and payment records. While borrowers must disclose their personal information, there is no requirement for lenders to do so. However, the ID number of all borrowers are available. Therefore, if a lender has posted a loan listing as a borrower in Renrendai, we can match it according to its ID. In this way, we are able to obtain personal information on a subset of lenders. The original data include 744,853 loan listings and 13,295,938 bidding records. To exploit the personal information fully, we consider only the listings for which both lenders and borrowers are identifiable. We drop the listings for which borrower or lenders are not identifiable from our empirical analysis. As a result, each node in our network represents an individual who is a lender or a borrower. To prevent estimation biases, we drop loan requests with incomplete information or with institutional guarantees. We also remove the listings for which the borrowers and the lenders are from Hong Kong, China; Macau, China; and Taipei, China, in consideration of their different business and legal environments. The organization of the resulting database can be either from the lenders' perspective, with one record for each listing for which a lender bids in a given year, or from the borrowers' perspective, with one record for each loan that a borrower posts in a given year. As a result, on the lenders' side, we obtain 273,508 bidding records for 116,460 loan listings from 2,723 individual lenders; on the borrowers' side, we obtain 44,481 loan listings from 17,585 individual borrowers, of which 23,365 were successful in obtaining funded while the remaining 21,116 did not receive funding. Among all the listings that successfully obtained funding, 3,618 defaulted.

3.2 Measurement of Key Variables

3.2.1 Network Centrality Measures

Mathematically, social network models belong to the class of graphical models. We can define a graphical model as a graph $G=(V, W)$, where V is a set of vertices (nodes) and $W=V \times V$ is a set of weights (edges or links) between all the vertices.

In a graphical Markov model (Lauritzen 1996), the weight set specializes to an edge set E , which describes whether any pair of vertices (i, j) is connected ($(i, j) \in E$) or not ($(i, j) \notin E$). An adjacency matrix, A , can fully specify a graphical Markov model. The adjacency matrix A of a vertex set V is the $|V| \times |V|$ matrix in which the entries are $a_{ij}=1$ if $(i, j) \in E$ and 0 otherwise.

From a statistical viewpoint, each vertex $v \in V$ in a graphical Markov model can be associated with a random variable X_v . When the vector of random variables $(X_v; v \in V)$ follows a multivariate Gaussian distribution, the model becomes a graphical Gaussian model, characterized by a correlation matrix R that we can use to derive the adjacency matrix. This is because the following equivalence holds:

$$(i, j) \notin E \iff (R^{-1})_{ij} = 0,$$

which states that a missing edge between vertex i and vertex j in the graph is equivalent to the partial correlation between variables X_i and X_j being equal to zero. Building on the previous equivalence, a graphical Gaussian model is able to learn from the data structure of a graph (the adjacency matrix) and, therefore, the dependence structure between the associated random variables. In particular, the model can retain an edge if the corresponding partial correlation is significantly different from zero.

In a social network analysis model (Barabási 2016), the set W is a set of weights, which usually connect each variable with all the others. From a statistical viewpoint, each vertex $v \in V$ in a network analysis model is associated with a statistical unit, and each weight

describes an observed relationship between a pair of units, such as a number of goods or a financial amount. While the adjacency matrix in a graphical Markov model is symmetric, the weight matrix does not need to be so. For instance, in interbank lending, which is one of the main applications of network analysis in the financial domain, the weights are financial transactions, with w_{ij} indicating how much i lends to j and w_{ji} indicating how much j lends to i . The aim of a network analysis model is not to learn from the data structure of a graph but, rather, to summarize a complex structure, which a graph describes in terms of summary measures, known as centrality measures. In the following, we describe the centrality measures that we believe to be relevant to our context.

Degree Centrality

Degree centrality counts the number of connections that a node has (Freeman 1978). In a directional graph, we can distinguish degree centrality into out-degree centrality and in-degree centrality. The former quantifies the tendency of nodes to “export” and the latter quantifies the tendency to “import.” Let l_{ij} (l_{ji}) indicate a connection from node i (j) to node j (i). We can define the out-degree centrality and in-degree centrality of node i , which $C_{o,i}$ and $C_{I,i}$ denote, respectively, as follows:

$$C_{o,i} = \sum_{j=1, j \neq i}^N \frac{l_{ij}}{N-1}$$

$$C_{I,i} = \sum_{j=1, j \neq i}^N \frac{l_{ji}}{N-1}.$$

Closeness

Closeness centrality focuses on how close a node is to all the other nodes; therefore, it measures how fast a given node i in a network can reach other nodes. If we indicate with d_{ij} the length of the minimal path between any two nodes, in terms of the number of edges connecting them, we can define the closeness centrality of a node as:

$$CC_i = \frac{N-1}{\sum_{j=1, j \neq i}^N d_{ij}}.$$

Betweenness

Betweenness centrality captures a node’s potential to control the information and resources that any other pair of nodes exchange. Let g_{jk} indicate the number of paths between a given pair (j, k) and $g_{jk}(i)$ the number of paths between j and k that contain i . We can define the betweenness centrality of node i , which we denote as BC_i , as follows:

$$BC_i = \frac{2 \sum_j^n \sum_k^n \frac{g_{jk}(i)}{g_{jk}}}{N^2 - 3N + 2}, j \neq k \neq i$$

Eigenvector Centrality

Eigenvector centrality reflects the idea that the centrality of a node depends not only on the number of its linked nodes but also on the centrality of the linked nodes. Unlike degree centrality, which weights every connection equally, eigenvector centrality weights contacts according to their centrality. It is also possible to see eigenvector centrality as a weighted sum not only of direct connections but also of indirect connections of any

length, thus potentially taking into account the entire pattern of the network (Bonacich 1972).

In matrix notation, if we denote the eigenvector centralities of a set of nodes with $EC=(EC_1, EC_2, \dots, EC_n)$ and the weight matrix with W , we can find the eigenvector centralities by solving the equation $W*EC=\lambda*EC$, which gives the eigenvalues of W , the corresponding eigenvectors of which are the centrality measures.

For interpretation purposes, we normalize all the above centrality measures in our analysis, dividing each centrality value by its theoretical maximum. Accordingly, all the measures can take values in the $[0,1]$ interval.

3.2.2 Measurement of the Lending Activity

We construct three variables to reflect the bidding activities of the lenders on the platform. We can define the first of them, *MONEY*, as the amount of money that a lender invests in each loan listing. At Renrendai, each lender can bid a minimum of 50 Chinese yuan or a multiple. The second variable is *LENDTIME*, measuring, for each loan, the percentage of the requested amount for which the borrower receives funding when a lender bids. It measures the timing that a lender involves in the bidding process. The larger the *LENDTIME*, the more likely a lender is to follow herding behavior for investment in the corresponding loan listing. According to Zhang and Liu (2012), many lenders may prefer a well-funded listing just because it is more likely to materialize into an actual loan. The third variable is *FSTB_M*, which we define as the time interval between a borrower posting a loan listing on the platform and the listing receiving a lender's bid. Every loan listing on Renrendai stays for at most seven days on the website, during which the lenders can invest in it until it receives complete funding. To avoid the influence of outliers, we winsorize *MONEY* and *FSTB_M* at the top and the bottom 1% percentile of their respective distributions. We also winsorize *LENDTIME* at the top 1% percentile.

3.2.3 Measurement of the Borrowing Activity

We use three variables to measure the borrowing activities of borrowers on the platform. The first variable, *INTEREST*, is the interest rate that a borrower offers for the loan application. At Renrendai, the lenders can only bid for an amount, not an interest rate. The second variable, *SUCCESS*, is a binary variable that is equal to one when the listed loan receives funding and zero otherwise. The third variable, *DEFAULT*, is also a binary variable that is equal to one when a loan listing is delinquent or defaults and zero otherwise.

In addition, we include two categories of control variables in the regressions. The first category is a lender's (or borrower's) personal characteristics, such as education, income, age, marital status, length of work experience, and so on. The second is the information about a loan listing, such as the maturity, interest rate, and borrowing amount. Table 1 summarizes the definition of all the variables that this paper uses.

Table 1: Variables and Definitions

Variable Name	Definition
<i>MONEY</i>	The amount that a lender invests in each investment.
<i>LENDTIME</i>	The percentage of the amount that the listing requests when lenders invest.
<i>FSTB_M</i>	The time interval between the loan listing being posted and the lender's investment.
<i>SUCCESS</i>	One if a loan listing is fully funded and zero otherwise.
<i>DEFAULT</i>	One if the funded loan has defaulted and zero otherwise.
<i>INTEREST</i>	The rate that the borrower pays on the loan.
<i>Indegree\Outdegree</i>	Proportional to the number of other nodes to which a node is linked—the number of links divided by (n-1). For a directed network, indegree is a count of the number of ties directed to the node and outdegree is the number of ties that the node directs to others.
<i>Betweenness</i>	The extent to which a particular point lies “between” other points in the graph; how many shortest paths (geodesics) is it on? A measure of brokerage or gatekeeping.
<i>Eigenvector</i>	The eigenvector centrality measures the connectivity between the borrower (the lender) and many other “central” lenders (the borrower). Its size depends on the centrality of the main contact objects of the node.
<i>Closeness</i>	The sum of geodesic distances (shortest paths) to all other points in the graph. Divide by (n-1), then invert.
<i>AMOUNT</i>	Loan amount requested by the borrower.
<i>MONTHS</i>	Loan term requested by the borrower.
<i>CREDIT</i>	Credit grade of the borrower at the time of the creation of the listing. The credit grade takes values between one (high risk) and seven (AA).
<i>House</i>	One if the borrower has a house and zero otherwise.
<i>Car</i>	One if the borrower has a car and zero otherwise.
<i>AGE</i>	Age of the borrower in years, winsorized at the 0.1 nd .
<i>EDUCATION</i>	Education level of the borrower: one=middle/high school, two=3-year college, three=4-year college, four=graduate school.
<i>INCOME</i>	Monthly income of the borrowers: one=less than 1,000, two=1,001–2,000, three=2,001–5,000, four=5,001–10,000, five=10,001–20,000, six=20,001–50,000, seven=more than 50,000.
<i>WORKTIME</i>	Borrowers' work experience: one=less than 1 year, two=1–3 years, three=3–5 years, four=more than 5 years.
<i>Married</i>	One if the borrower is married and zero otherwise.
<i>N_Length</i>	The number of characters in the borrower's nickname.
<i>T_Length</i>	The number of characters in a loan title.
<i>D_Length</i>	The number of characters in a loan description.
<i>Purpose</i>	The usage of the fund described by the borrowers, including short-term turnover, personal consumption, car loans, housing loans, wedding planning, educational training, other loans, investment, medical expenditure, and decorating loans.
<i>Industry</i>	The industry that a borrower is working for, including IT, restaurant/hotel, real estate, public utilities, public welfare organizations, computer systems, construction, transportation, education/training, finance, law, retail/wholesale, media/advertising, energy, agriculture, other, sports/arts, entertainment, medical/sanitation/health care, government agencies, and manufacturing.
<i>Region</i>	The area in which the borrower is located.

3.3 Summary Statistics

Table 2 presents the summary statistics for the key variables.

Table 2: Summary Statistics

Panel A		Loan Characteristics			
Variable	Mean	SD	Min.	Max.	N
SUCCESS	0.525	0.499	0	1	44,481
DEFAULT	0.155	0.362	0	1	23,365
Indegree	0.035	0.131	0	2.042	44,481
Outdegree	0.011	0.107	0	1.598	44,481
Betweenness	0.001	0.013	0	0.186	44,481
Eigenvector	0.005	0.018	0	0.218	44,481
Closeness	0.004	0.007	0	0.075	44,481
INTEREST	13.08	2.600	3	24.40	44,481
AMOUNT	45,000	76,000	3,000	500,000	44,481
MONTHS	13.900	9.279	1	36	44,481
CREDIT	1.946	1.406	1	7	44,481
AGE	34.640	6.629	25	54	44,481
HOUSE	0.562	0.496	0	1	44,481
CAR	0.384	0.486	0	1	44,481
D_length	111.600	94.960	10	973	44,481
T_length	15.230	7.866	1	59	44,481
N_length	8.315	2.749	2	28	44,481
EDUCATION	2.162	0.819	1	4	44,481
Edu=HighSchool	0.234	0.423	0	1	44,481
Edu=JuniorCollege	0.404	0.491	0	1	44,481
Edu=Bachelor	0.328	0.470	0	1	44,481
Edu=Postgraduate	0.034	0.180	0	1	44,481
WORKTIME	2.779	1.017	1	4	44,481
Worktime<=1year	0.095	0.293	0	1	44,481
Worktime=1~3year	0.367	0.482	0	1	44,481
Worktime=3~5year	0.202	0.401	0	1	44,481
Worktime>=5year	0.336	0.472	0	1	44,481
INCOME	4.226	1.325	1	7	44,481
Income<=1000	0.003	0.052	0	1	44,481
Income=1~2000	0.011	0.105	0	1	44,481
Income=2~5000	0.361	0.480	0	1	44,481
Income=5~10000	0.297	0.457	0	1	44,481
Income=1~20000	0.129	0.336	0	1	44,481
Income=2~50000	0.108	0.310	0	1	44,481
Income>50000	0.091	0.287	0	1	44,481
Purpose=Shortturnover	0.480	0.500	0	1	44,481
Purpose=Consumption	0.105	0.306	0	1	44,481
Purpose=Buycar	0.048	0.214	0	1	44,481
Purpose=Buyhouse	0.029	0.167	0	1	44,481
Purpose=Wedding	0.017	0.129	0	1	44,481
Purpose=Education	0.012	0.108	0	1	44,481
Purpose=Other	0.044	0.205	0	1	44,481
Purpose=Startbusiness	0.104	0.306	0	1	44,481
Purpose=Medical	0.002	0.043	0	1	44,481
Purpose=Decoration	0.159	0.366	0	1	44,481
Ind=IT	0.090	0.287	0	1	44,481

continued on next page

Table 2 *continued*

Panel A		Loan Characteristics			
Variable	Mean	SD	Min.	Max.	N
Ind=Restauranthotel	0.022	0.147	0	1	44,481
Ind=Realestate	0.027	0.163	0	1	44,481
Ind=Publicutilities	0.020	0.141	0	1	44,481
Ind=Commonweal	0.001	0.034	0	1	44,481
Ind=Computer	0.000	0.016	0	1	44,481
Ind=Construction	0.040	0.197	0	1	44,481
Ind=Transportation	0.045	0.208	0	1	44,481
Ind=Education	0.048	0.214	0	1	44,481
Ind=FinanceLaw	0.050	0.217	0	1	44,481
Ind=Retail	0.140	0.347	0	1	44,481
Ind=Mediaadvertising	0.024	0.153	0	1	44,481
Ind=Energy	0.046	0.209	0	1	44,481
Ind=Agriculture	0.015	0.121	0	1	44,481
Ind=Other	0.075	0.264	0	1	44,481
Ind=Sportart	0.004	0.063	0	1	44,481
Ind=Medical	0.033	0.179	0	1	44,481
Ind=Entertainment	0.012	0.111	0	1	44,481
Ind=Governmentoffice	0.094	0.291	0	1	44,481
Ind=Manufacturing	0.213	0.409	0	1	44,481
Region=East	0.602	0.490	0	1	44,481
Region=West	0.139	0.346	0	1	44,481
Region=Northeast	0.053	0.223	0	1	44,481
Region=Middle	0.206	0.405	0	1	44,481
YEAR=2011	0.081	0.272	0	1	44,481
YEAR=2012	0.098	0.297	0	1	44,481
YEAR=2013	0.119	0.324	0	1	44,481
YEAR=2014	0.368	0.482	0	1	44,481
YEAR=2015	0.335	0.472	0	1	44,481
Panel B		Investment Behavior			
Variable	Mean	SD	Min.	Max.	N
MONEY	676.6	1490	50	10,000	273,508
LENDTIME	0.031	0.08	0	1	273,508
FSTB_M	1,735	3,483	0.167	16,000	273,508
L_EDUCATION	2.383	0.861	1	4	2,723
L_WORKTIME	2.662	1.064	1	4	2,723
L_INCOME	4.023	1.215	1	7	2,723
L_AGE	34.096	5.886	25	57	2,723
L_HOUSE	0.581	0.493	0	1	2,723
L_CAR	0.388	0.487	0	1	2,723
L_Married	0.622	0.485	0	1	2,723
YEAR=2010	0.004	0.064	0	1	273,508
YEAR=2011	0.134	0.341	0	1	273,508
YEAR=2012	0.0980	0.298	0	1	273,508
YEAR=2013	0.161	0.368	0	1	273,508
YEAR=2014	0.209	0.406	0	1	273,508
YEAR=2015	0.280	0.449	0	1	273,508
YEAR=2016	0.114	0.317	0	1	273,508

continued on next page

Table 2 *continued*

Panel C		Difference Test: By Median			
Variable	Indegree<Median	Mean	Indegree>Median	Mean	Diff.
MONEY	200,297	707.9	73,211	591.1	116.72***
LENDTIME	200297	0.02	73,211	0.06	-0.04***
FSTB_M	200,297	1,738	73,211	1,725	13.14
INTEREST	22,247	12.82	22,234	13.34	-0.52***
SUCCESS	22,247	0.460	22,234	0.590	-0.13***
DEFAULT	10,218	0.230	13,147	0.100	0.13***
	Outdegree<Median	Mean	Outdegree>Median	Mean	Diff.
MONEY	136,755	731.0	136,753	622.3	108.67***
LENDTIME	136,755	0.01	136,753	0.05	-0.03***
FSTB_M	136,755	1,902	136,753	1,568	334.29***
INTEREST	37,946	13.15	6,535	12.66	0.50***
SUCCESS	37,946	0.560	6,535	0.350	0.20***
DEFAULT	21,074	0.170	2,291	0.02	0.14***
	Betweenness<Median	Mean	Betweenness>Median	Mean	Diff.
MONEY	223,436	697.9	50,072	581.6	116.32***
LENDTIME	223,436	0.0200	50,072	0.0800	-0.06***
FSTB_M	223,436	1,767	50,072	1,592	175.34***
INTEREST	42,070	13.13	2,411	12.27	0.86***
SUCCESS	42,070	0.510	2,411	0.740	-0.23***
DEFAULT	21,578	0.170	1,787	0.0200	0.15***
	Eigenvector<Median	Mean	Eigenvector>Median	Mean	Diff.
MONEY	137,271	724.6	136,237	628.2	96.44***
LENDTIME	137,271	0.0100	136,237	0.0500	-0.03***
FSTB_M	137,271	1,916	136,237	1,553	362.92***
INTEREST	22,243	12.97	22,238	13.19	-0.21***
SUCCESS	22,243	0.480	22,238	0.570	-0.09***
DEFAULT	10,669	0.230	12,696	0.0900	0.14***
	Closeness<Median	Mean	Closeness>Median	Mean	Diff.
MONEY	137,253	735.4	136,255	617.4	118.04***
LENDTIME	137,253	0.0200	136,255	0.0500	-0.03***
FSTB_M	137,253	1,896	136,255	1,573	322.51***
INTEREST	22,242	12.96	22,239	13.20	-0.24***
SUCCESS	22,242	0.500	22,239	0.560	-0.06***
DEFAULT	11,014	0.240	12,351	0.0800	0.17***

Note: Panel C reports the mean difference test results in the column "Diff." *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A in Table 2 presents the summary statistics of loan listings and borrowers' personal characteristics. Note that the funding success rate of the loan listings is about 52%, a relatively low percentage, regardless of the average interest of 13.08%. Among the listings that successfully obtained funding, the default rate is high, about 15%. In terms of centrality measures, indegree centrality is much higher than outdegree centrality, reflecting the fact that each borrower is associated with more than one lender. The average loan amount is approximately 45,000 Chinese yuan, indicating the market's demand for small loans. The mean age of borrowers is 34 years, reflecting the fact that peer-to-peer lending is more attractive to young people. In terms of

socio-economic status, 65.8% of the borrowers have an income in the range of 2,001–10,000 Chinese yuan per month; 76.6% of the borrowers have an educational level corresponding to a 3-year college degree or above. Furthermore, 56.2% of the borrowers own a house; 38.4% own a car; and 91% have been working for more than 1 year. In summary, the borrowers using the Chinese P2P platform are young, well educated, and have medium income and wealth and limited work experience.

Panel B in Table 2 presents the summary statistics of lending behavior and lenders' personal characteristics. Panel A shows that the average amount of each investment is 676 Chinese yuan, much higher than the minimum investment amount of 50 Chinese yuan that Renrendai sets. The average *LENDTIME* is 3%, meaning that the investors on average bid for a loan listing when only 3% of the requested amount has attracted bids. This implies that lenders are willing to invest in the early stage of loan listing rather than following herding behavior. The average time interval between the loan listing being posted and it receiving a new bid is 1,735 minutes (28.9 hours). This means that lenders need time to acquire basic information and think before making investment decisions. In terms of lenders' characteristics, we find that the lenders who participate in the platform are young and well educated with a medium income. In more detail, the average age of lenders is less than 35 years; their average level of education is above the 3-year college level; and the average lender's monthly income ranges from 5,000 to 10,000 Chinese yuan. Consistent with their young age, the work experience of most lenders is limited, with an average of about 1 to 3 years. Among all the lenders, 62.2% are married, 58.1% own a house, and 38.8% own a car.

Panel C in Table 2 statistically compares the lending and borrowing activities conditionally using various centrality measures. We capture the lending activities through the three variables of *MONEY*, *LENDTIME*, and *FSTB_M* and the borrowing activities through the three variables of *INTEREST*, *SUCCESS*, and *DEFAULT*. We divide the whole sample into two groups according to whether a loan listing's centrality measurements are above or below the medium value and then implement the difference tests on various lending and borrowing activities. Taking the measurement of *Indegree*, for instance, there are 200,297 listings below the median and 73,211 listings above the median. The lenders whose centrality measures are above the median on average invest around 100 Chinese yuan more than their peers whose centrality is below the median.

3.4 Network-Based Regression

To quantify the impact of network centrality on market outcomes, we propose to incorporate various measurements of network centrality into a linear regression model as additional explanatory variables. More formally, we assume that, for each node i in a network (which can be a borrower or a lender), the dependent variable y_i is linearly related to the measurement of a set of J control variables (x_{ij}) and a set of K centrality variables (g_{ik}) as follows:

$$y_i = \alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k g_{ik} + \mu_i + \sigma_t + \varepsilon_i, \quad (1)$$

where the available data allow the estimation of the coefficients β_j and γ_k , typically with the ordinary least square method; μ_i and σ_t represent loan and time fixed effects, respectively; and ε_i is the error term.

We can extend a similar approach to the contexts in which the dependent variables are binary, such as the default of a loan. For example, in the context of P2P lending, Duarte, Siegel, and Young (2012), Iyer et al. (2016), Lin and Viswanathan (2016), and Chen, Huang, and Ye (2018) used logistic regression. These authors classified P2P borrowers

into two groups, characterized by different payment histories of the loans that received funding through the platform: 0=active (all loans have been paid on time); 1=default (at least one loan has not been paid on time). A logistic regression model estimates the probability that a borrower will default, using data on a set of borrower-specific variables. More formally:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_j \beta_j x_{ij} + \mu_i + \sigma_t + \varepsilon_i, \quad (2)$$

where for each borrower $i=(1, \dots, I)$; p_i is the probability of default; $x_i=(x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of borrower-specific explanatory variables; the intercept parameter α and the regression coefficients β_j , for $j=1, \dots, J$, are unknown and we must estimate them from the available data; μ_i and σ_t represent loan and time fixed effects, respectively; and ε_i is the error term.

From the previous expression, we can obtain the probability of default for each borrower as follows:

$$p_i = \frac{1}{1 + e^{\alpha + \sum_j \beta_j x_{ij}}}$$

the credit score of i , whose default status we can predict as one or zero depending on whether p_i exceeds a set threshold θ . Common choices for the threshold are $\theta=0.5$ and $\theta=d/I$, with d being the observed number of defaults.

We propose to embed the network centrality measure as a key explanatory variable into a logistic regression model. More formally, our proposed network-based scoring model takes the following form:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k g_{ik} + \mu_i + \sigma_t + \varepsilon_i, \quad (3)$$

where p_i is the probability of default for borrower i ; $x_i=(x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of borrower-specific explanatory variables, g_{ik} is the degree of centrality measure for borrower i under the measurement k ; and the intercept parameter α and the regression coefficients β_j and γ_k , for $j=1, \dots, J$ and $k=1, \dots, K$, require estimation from the available data. It follows that we can obtain the probability of default as follows:

$$p_i = \frac{1}{1 + e^{\alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k g_{ik}}}.$$

To summarize, we “augment” both linear and logistic regression models by means of the proposed centrality measures, improving their explanatory power. By doing so, we combine the explainability that a “simple” regression model provides with the accuracy of a “complex” network model.

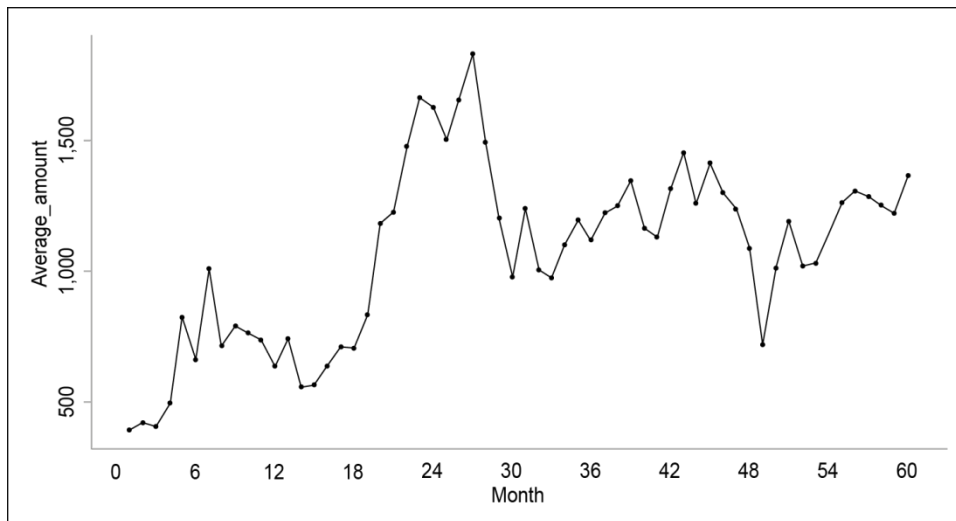
4. EMPIRICAL ANALYSIS

In this section, we empirically verify, using network-based regression models, our hypotheses that network centrality influences both investment behavior and loan outcomes.

4.1 Evolution of the P2P Network

We first present the estimation results of the network models from which we obtain the centrality measures.

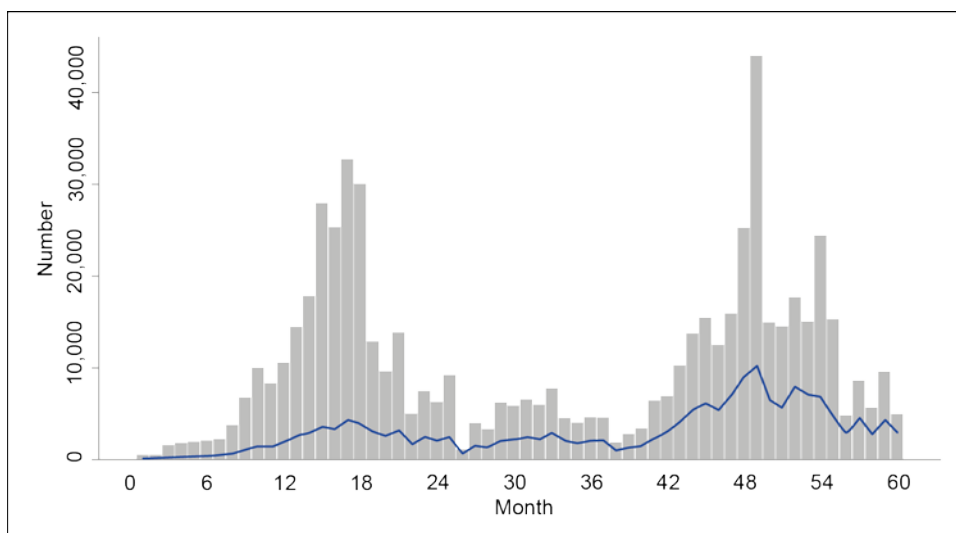
Figure 1: Average Amount of Lending



Note: The figure represents the time evolution of the average amount of lending, which we calculated using data from the subsequent months in the period 2011–2015.

Figure 1 shows the evolution of the average amount of lending over time. It was small in 2011 but increased dramatically in 2012 and finally decreased and stabilized at around the amount of 1250 Chinese yuan thereafter. This reflects the initial enthusiasm of Chinese investors for P2P lending, which, after the PRC government tightened the regulation on the P2P lending market, gradually returned to the normal level.

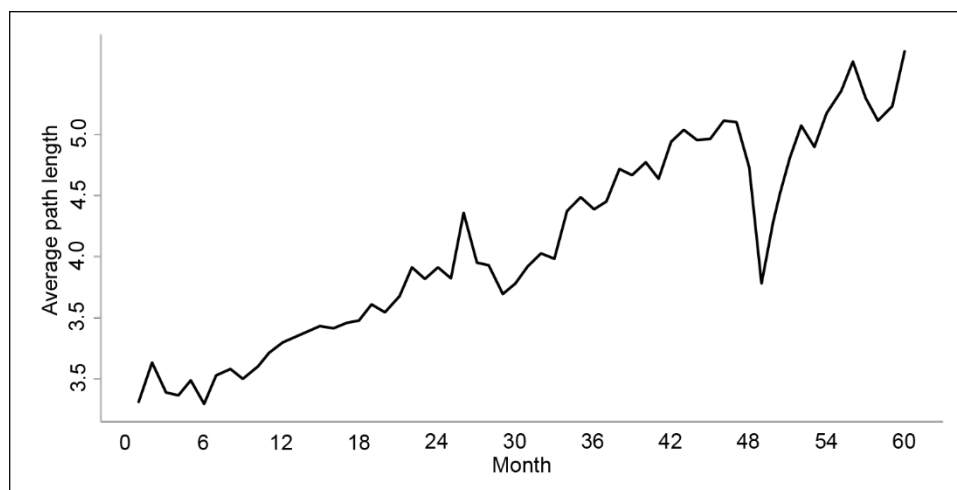
Figure 2: Edge and Node Numbers



Note: The figure represents the time evolution of the number of nodes (blue) and of the number of edges (grey) between them according to the networks that we estimated using data from the subsequent months in the period 2011–2015. Blue line and gray bar indicate the changes of node number and edge number respectively.

Figure 2 shows the evolution of the number of nodes (borrowers and lenders) and of the links between them over time. Note that, while the number of nodes has increased (especially during 2014 and 2015), the number of links has not grown correspondingly, implying that the network has become sparser. From an economic viewpoint, this means that fewer lenders, who are not inclined to diversify their portfolio, have co-invested in each loan listing. In addition, initially some lenders could borrow money from the platform to lend to other borrowers on the same platform. However, due to the tightened regulation, it has become harder and harder to borrow and reinvest through Renrendai.

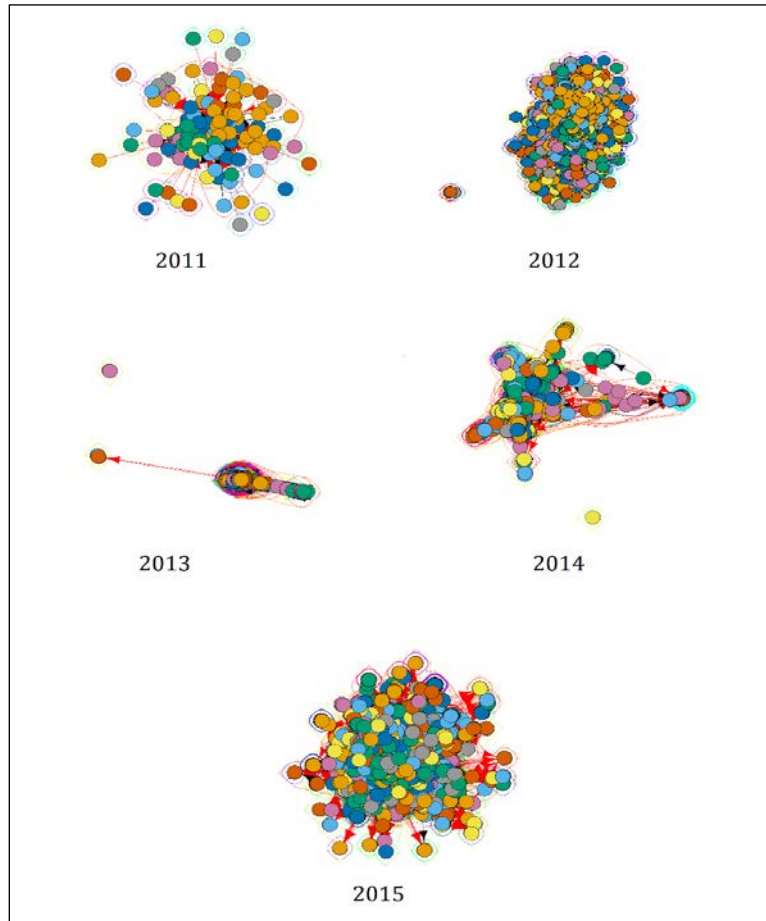
Figure 3: Average Path Length



Note: The figure represents the time evolution of the average path length according to the networks that we estimated using data from the subsequent months in the period 2011–2015.

Figure 3 shows the evolution of the average path length. Consistent with Figure 2, the average path length has increased over time. It shows that, from 2011 to 2015, the number of investors and borrowers continued to increase, although the amount invested (Figure 1) and the diversification (Figure 2) stabilized. This may also be due to the continuous entrance of small and inexperienced participants.

For comparison, Figure 4 plots one network for each year from 2011 to 2015. For each yearly network, we apply the random-walk method that Pons and Latapy (2006) proposed to cluster the borrowers into community groups. This helps us to visualize how the nodes (borrowers and lenders) are integrated into the same community and how a “community map” evolves over time. The comparison of the five yearly maps from 2011 to 2015 shows that the number of communities has increased considerably. Starting from 602 in 2011, the number of communities grew to 1070 in 2012, 3284 in 2013, and 6602 in 2014; finally, the largest number of communities of 8492 covering 7827 borrowers occurred in 2015. In the years 2012, 2013, and 2014, some isolated communities emerged, meaning that the overall network did not integrate a few outlying borrowers and lenders well. This may be because of the growing number of new entrants into the platform, which increases their heterogeneity and distance. However, the further development of the network incorporated these outliers into the core network in 2015.

Figure 4: Time Evolution of the Renrendai Network

Note: The figure represents, in every year, each borrower and lender active on the platform with a node. The links between nodes describe transactions between each pair of nodes. Each node is colored according to the community to which the community detection algorithm attributes it.

4.2 Network and Lending Behavior

We examine the effect of network centrality on lending behavior with the following regression model:

$$Lending_i = \beta_0 + \beta_1 Network_i + \beta_2 Control_i + \mu_i + \sigma_t + \varepsilon_i, \quad (4)$$

where the dependent variable $Lending_i$ measures the behavior of lender i with three different variables, $MONEY_i$, $LENDTIME_i$, and $FSTB_M_i$, respectively. The main explanatory variable of interest is $Network_i$, a vector of network centrality measures for lender i , including $Indegree_i$, $Outegree_i$, $Betweenness_i$, $Eigenvector_i$, and $Closeness_i$. $Control_i$ is a set of control variables reflecting lenders' personal characteristics. The parameters μ_i and σ_t represent loan and time fixed effects, respectively. Finally, ε_i is the random disturbance term.

Table 3: Linear Regression of Investment Amount

	(1)	(2)	(3)	(4)	(5)
	MONEY	MONEY	MONEY	MONEY	MONEY
Indegree	74.5570*** (4.04)				
Outdegree		163.0385*** (8.08)			
Betweenness			816.3230*** (4.71)		
Eigenvector				2.0e+03*** (9.34)	
Closeness					-3.7e+03*** (-13.68)
L_EDUCATION	32.6798*** (6.99)	26.9420*** (5.74)	32.4160*** (6.94)	29.3311*** (6.26)	42.7172*** (9.12)
L_WORKTIME	-2.0667 (-0.52)	1.4425 (0.37)	-1.5558 (-0.39)	-1.9152 (-0.48)	-2.9530 (-0.75)
L_INCOME	70.4721*** (19.55)	68.9037*** (19.10)	71.2527*** (19.85)	70.5991*** (19.65)	69.6537*** (19.45)
L_AGE	13.7753*** (19.08)	13.6978*** (18.97)	13.7634*** (19.06)	13.8773*** (19.22)	13.8288*** (19.16)
L_HOUSE	39.2646*** (4.00)	43.8978*** (4.49)	41.6515*** (4.26)	35.7611*** (3.65)	46.2697*** (4.73)
L_CAR	73.8753*** (8.81)	68.6631*** (8.11)	73.5305*** (8.74)	67.1158*** (7.97)	96.0841*** (11.29)
L_Married	-94.2651*** (-8.98)	-1.0e+02*** (-9.99)	-98.4834*** (-9.49)	-92.4595*** (-8.88)	-1.0e+02*** (-10.09)
_cons	-1.3e+02 (-1.24)	64.0789 (0.62)	57.9334 (0.56)	67.8931 (0.66)	49.5794 (0.48)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
F	33.6386	34.9708	33.8062	35.5191	35.9883
r2_a	0.3563	0.3566	0.3563	0.3569	0.3572

Note: (1) This table reports the OLS regression results on the investment amount. The dependent variable is *MONEY*, the amount that a lender invests in each investment. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *L_EDUCATION*—measuring the education level of a lender; *L_WORKTIME*—a lender's work experience measured in years; *L_INCOME*—the monthly income of a lender; *L_AGE*—the age of a lender expressed in years; *L_HOUSE*—a dummy variable taking the value one if the lender is a homeowner and zero otherwise; *L_CAR*—a dummy variable taking the value one if a lender owns a car and zero otherwise; and *L_Married*—a dummy variable taking the value one if a lender is married and zero otherwise; all the regressions include loan fixed effects, year fixed effects, day-of-week fixed effects, and hour-of-day fixed effects.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the T-statistics in parentheses. N is the number of observations. F is the F statistic. r2_a is the adjusted R-square.

Table 3 reports the ordinary least squares (OLS) estimation results for *MONEY*, the amount of money that a lender invests in a loan. The network centrality measures exhibit strong explanatory power for the amount invested in a loan. The coefficients for the variables *Indegree*, *Betweenness*, and *Eigenvector* are all positive and significant at the 1% statistical level. These results indicate that lenders' social network connections significantly enhance their investment amount. From an economic viewpoint, compared with those lenders with fewer connections, well-connected lenders have more channels through which to gain information, more accurate knowledge of borrowers' risks, and

hence no need to make smaller investments to diversify their portfolios. Furthermore, a large number of links implies an information advantage, which make the lenders more confident and more inclined to take risk (Faleye, Kovacs, and Venkateswaran 2014). From Table 3, we also note that the estimates for most of the control variables are consistent with our expectations. The investors with a higher level of education, income, and wealth tend to invest a larger amount in a loan listing.

Table 4: Linear Regression Result of Investment Timing

	(1)	(2)	(3)	(4)	(5)
	LENDTIME	LENDTIME	LENDTIME	LENDTIME	LENDTIME
Indegree	-0.0201*** (-13.79)				
Outdegree		-0.0098*** (-6.36)			
Betweenness			-0.1595*** (-10.02)		
Eigenvector				-0.1050*** (-5.45)	
Closeness					-0.0105 (-0.65)
L_EDUCATION	-0.0011*** (-5.03)	-0.0009*** (-4.01)	-0.0011*** (-5.01)	-0.0011*** (-4.86)	-0.0013*** (-5.82)
L_WORKTIME	0.0015*** (7.41)	0.0012*** (5.76)	0.0014*** (6.68)	0.0014*** (6.70)	0.0014*** (6.52)
L_INCOME	-0.0026*** (-14.70)	-0.0028*** (-15.91)	-0.0029*** (-16.21)	-0.0030*** (-16.54)	-0.0030*** (-17.11)
L_AGE	-0.0004*** (-12.60)	-0.0004*** (-11.96)	-0.0004*** (-12.35)	-0.0004*** (-12.29)	-0.0004*** (-11.92)
L_HOUSE	-0.0023*** (-4.36)	-0.0042*** (-7.80)	-0.0033*** (-6.28)	-0.0038*** (-7.05)	-0.0044*** (-8.09)
L_CAR	-0.0001 (-0.33)	-0.0007 (-1.49)	-0.0004 (-0.83)	-0.0006 (-1.43)	-0.0012*** (-2.59)
L_Married	0.0043*** (7.80)	0.0076*** (13.25)	0.0061*** (10.81)	0.0070*** (12.23)	0.0078*** (13.60)
_cons	0.0369*** (4.93)	0.0953*** (19.10)	0.0932*** (18.65)	0.0953*** (19.08)	0.0968*** (19.45)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
F	27.8022	24.9796	26.5491	25.1179	24.6239
r2_a	0.3708	0.3678	0.3694	0.3680	0.3672

Note: (1) This table reports the OLS regression results for investment timing. The dependent variable is *LENDTIME*, the percentage of the amount that the listing requests when lenders invest. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *L_EDUCATION*—measuring the education level of a lender; *L_WORKTIME*—a lender's work experience measured in years; *L_INCOME*—the monthly income of a lender; *L_AGE*—the age of a lender expressed in years; *L_HOUSE*—a dummy variable taking the value one if the lender is a homeowner and zero otherwise; *L_CAR*—a dummy variable taking the value one if a lender owns a car and zero otherwise; and *L_Married*—a dummy variable taking the value one if a lender is married and zero otherwise; all the regressions include loan fixed effects, year fixed effects, day-of-week fixed effects, and hour-of-day fixed effects.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the T-statistics in parentheses. N is the number of observations. F is the F statistic. r2_a is the adjusted R-square.

We further assess the impacts of network centrality on investors' timing of a bid (*LENDTIME*) and present the OLS estimation results in Table 4. Columns (1)–(5) show

that the coefficients of *Indegree*, *Outdegree*, *Betweenness*, and *Eigenvector* are all negative and significant at the 1% confidence level while the coefficient of *Closeness* is negative but not significant. Recall that *LENDTIME* gauges, for each loan, the percentage of the requested amount that has obtained funding when a lender bids. The smaller the value of *LENDTIME*, the more swiftly a lender makes the bids, signifying a higher level of trust in a loan listing. The obtained negative coefficients for various centrality measures indicate that network connectedness encourages the lenders to invest promptly after the posting of a listing. In a P2P lending market, lenders tend to invest at the later stage of the bidding process, so their investments can yield earnings as soon as possible. Our results essentially indicate that the information advantage that the social network linkage fosters spurs lenders to take the leading role in the bidding process rather than adopting herding behavior. From Table 4, we also note that, among the control variables, most of them are significant and negative, with the exception of work experience and marriage status. The investors with higher levels of education, income, and wealth tend to invest at the early stage of bidding.

Table 5 shows the OLS estimation results for *FSTB_M*, another variable reflecting investors' speed of bidding. Columns (1)–(5) show that the coefficients of *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness* are all negative and significant, either at the 5% or at the 1% confidence level, implying that the more central a lender's location in a network, the shorter the interval between the time when the lender invests in a loan listing and the time of posting of the loan listing. This further confirms that the higher the number of social network linkages a lender has, the faster is his or her investment action. This finding reinforces our assumption that the information advantage that a lender enjoys in a network enables him or her to make swift decisions. From Table 5, we also note that the control variables affect the lending performance in the same (positive) direction, with the exception of work time and marital status, which, as in the previous regression, show the opposite sign.

To summarize, the empirical evidence presented in this subsection indicates that network centrality improves the loan outcome by increasing the amount and the speed of the funding. In other words, we have empirically verified our **Hypothesis 1**.

4.3 Network and Borrowing Outcome

We now examine the effects of social networks on the borrowing outcome, which we measure using the borrowing rate, funding success, and default probability. We estimate the effect of network centrality on the borrowing rate through the following regression model:

$$Interest_j = \beta_0 + \beta_1 Network_j + \beta_2 Control_j + \mu_j + \sigma_t + \varepsilon_j, \quad (5)$$

where $Interest_j$ is the interest rate that borrower j offers; $Network_j$ is a vector of centrality measures for borrower j , including *Indegree_j*, *Outdegree_j*, *Betweenness_j*, *Eigenvector_j*, and *Closeness_j*. $Control_j$ is a set of control variables reflecting the loan and the borrower's personal characteristics. The parameters μ_i and σ_t represent loan and time fixed effects, respectively. Finally, ε_j is the error term. Table 6 reports the OLS estimation results.

Table 5: Linear Regression of the Investment Interval

	(1)	(2)	(3)	(4)	(5)
	FSTB_M	FSTB_M	FSTB_M	FSTB_M	FSTB_M
Indegree	-5.5e+02*** (-27.71)				
Outdegree		-2.7e+02*** (-18.49)			
Betweenness			-4.6e+03*** (-24.95)		
Eigenvector				-3.3e+03*** (-19.85)	
Closeness					-2.0e+03*** (-11.45)
L_EDUCATION	5.0234** (2.38)	10.7136*** (4.98)	5.3262** (2.51)	6.8763*** (3.22)	4.9741** (2.38)
L_WORKTIME	14.6432*** (7.33)	5.2427*** (2.58)	10.7042*** (5.30)	10.8666*** (5.39)	9.2486*** (4.55)
L_INCOME	-20.7130*** (-11.79)	-26.4505*** (-14.48)	-27.6234*** (-15.20)	-29.2304*** (-15.86)	-32.7620*** (-17.32)
L_AGE	-2.3402*** (-7.08)	-1.8085*** (-5.40)	-2.1518*** (-6.48)	-2.1147*** (-6.33)	-1.7523*** (-5.21)
L_HOUSE	-17.2484*** (-3.19)	-68.3907*** (-11.65)	-43.7832*** (-7.90)	-54.3858*** (-9.57)	-73.8319*** (-12.37)
L_CAR	-13.3696*** (-2.99)	-27.2610*** (-5.99)	-18.2736*** (-4.05)	-24.2738*** (-5.33)	-32.9168*** (-7.12)
L_Married	43.2207*** (7.76)	132.0515*** (19.02)	88.9701*** (14.55)	112.5833*** (17.31)	138.8808*** (19.63)
_cons	1.1e+03*** (8.97)	2.5e+03*** (30.80)	2.4e+03*** (30.70)	2.5e+03*** (30.94)	2.6e+03*** (31.21)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
F	43.9426	36.9802	41.3278	37.7440	36.0357
r2_a	0.9696	0.9684	0.9691	0.9686	0.9683

Note: (1) This table reports the OLS regression results for the investment interval. The dependent variable is *FSTB_M*, the time interval between the posting of the loan listing and the lender's investment. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *L_EDUCATION*—measuring the education level of a lender; *L_WORKTIME*—a lender's work experience measured in years; *L_INCOME*—the monthly income of a lender; *L_AGE*—the age of a lender expressed in years; *L_HOUSE*—a dummy variable taking the value one if the lender is a homeowner and zero otherwise; *L_CAR*—a dummy variable taking the value one if a lender owns a car and zero otherwise; and *L_Married*—a dummy variable taking the value one if a lender is married and zero otherwise; all the regressions include loan fixed effects, year fixed effects, day-of-week fixed effects, and hour-of-day fixed effects.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the T-statistics in parentheses. N is the number of observations. F is the F statistic. r2_a is the adjusted R-square.

Table 6: Linear Regression of the Loan Interest Rate

	(1)	(2)	(3)	(4)	(5)
	INTEREST	INTEREST	INTEREST	INTEREST	INTEREST
Indegree	-1.8136*** (-12.97)				
Outdegree		-3.1886*** (-29.29)			
Betweenness			-26.2935*** (-23.29)		
Eigenvector				-15.1831*** (-17.78)	
Closeness					-58.1213*** (-20.66)
InAMOUNT	0.1584*** (12.84)	0.1345*** (10.91)	0.1362*** (11.01)	0.1582*** (12.86)	0.1204*** (9.69)
MONTHS	0.0743*** (53.90)	0.0753*** (54.97)	0.0758*** (54.94)	0.0739*** (53.68)	0.0741*** (54.08)
CREDIT	-0.2794*** (-26.32)	-0.2698*** (-27.11)	-0.2669*** (-26.65)	-0.2697*** (-25.65)	-0.2308*** (-21.19)
T_length	0.0216*** (13.16)	0.0198*** (12.19)	0.0200*** (12.29)	0.0215*** (13.13)	0.0205*** (12.58)
N_length	-0.0191*** (-5.37)	-0.0203*** (-5.71)	-0.0177*** (-4.98)	-0.0197*** (-5.54)	-0.0216*** (-6.07)
D_length	0.0002 (1.24)	0.0002* (1.69)	0.0002 (1.52)	0.0002 (1.34)	0.0001 (0.99)
EDUCATION	-0.1379*** (-9.79)	-0.1363*** (-9.80)	-0.1412*** (-10.14)	-0.1378*** (-9.79)	-0.1368*** (-9.80)
WORKTIME	0.0249** (1.99)	0.0013 (0.11)	0.0023 (0.19)	0.0220* (1.76)	0.0130 (1.05)
INCOME	-0.0433*** (-3.76)	-0.0524*** (-4.55)	-0.0489*** (-4.25)	-0.0470*** (-4.08)	-0.0515*** (-4.45)
AGE	0.0033* (1.69)	0.0028 (1.41)	0.0028 (1.40)	0.0035* (1.77)	0.0031 (1.56)
HOUSE	-0.1085*** (-4.40)	-0.1024*** (-4.15)	-0.1018*** (-4.13)	-0.1045*** (-4.23)	-0.1015*** (-4.11)
CAR	-0.1792*** (-7.24)	-0.1595*** (-6.48)	-0.1668*** (-6.78)	-0.1739*** (-7.03)	-0.1678*** (-6.82)
Married	-0.1451*** (-5.68)	-0.1304*** (-5.12)	-0.1378*** (-5.40)	-0.1382*** (-5.41)	-0.1314*** (-5.14)
_cons	9.6592*** (66.75)	9.9639*** (69.17)	9.9151*** (68.79)	9.6660*** (67.01)	10.1171*** (70.05)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
F	429.9738	449.4826	440.4068	434.8225	439.4470
r2_a	0.2639	0.2723	0.2711	0.2657	0.2705

Note: (1) This table reports the OLS regression results for the interest rate. The dependent variable is *INTEREST*, the interest rate that the borrower pays on the loan. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *InAMOUNT*—the natural log of the loan amount (in RMB) that the borrower requests; *MONTHS*—the loan term (in months) that the borrower requests; *CREDIT*—the credit grade of the borrower at the time of creating the listing; *T_Length*—the number of characters in a loan title; *N_Length*—the number of characters in a borrower's nickname; *D_Length*—the number of characters in a loan description; *EDUCATION*—measuring the education level of a borrower; *WORKTIME*—a borrower's work experience measured by years; *INCOME*—the monthly income of a borrower; *AGE*—the age of a borrower expressed in years; *HOUSE*—a dummy variable taking the value one if the borrower is a homeowner and zero otherwise; *CAR*—a dummy variable taking the value one if a borrower owns a car and zero otherwise; *Married*—a dummy variable taking the value one if a borrower is married and zero otherwise; *Year*—a dummy controlling the year fixed effect; *Region*—a dummy variable reflecting the area in which a borrower is located; *Industry*—a dummy variable reflecting the industry that a borrower is working in; and *Purpose*—a dummy describing different purposes of borrowing.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the T-statistics in parentheses. N is the number of observations. F is the F statistic. r2_a is the adjusted R-square.

In Table 6, all the centrality parameters are negative and significant at the 5% or 10% confidence level, indicating that the higher the centrality, the lower the interest rate that a borrower can offer. In other words, the more central is the borrower, the higher is his or her bargaining power. This finding echoes several existing studies in the finance literature. For example, Engelberg, Gao, and Parsons (2013) demonstrated that CEOs with large networks earn more than those with small networks, while Hochberg, Ljungqvist, and Lu (2010) showed that strong networks among incumbent venture capitalists (VCs) in local markets improve their bargaining power over entrepreneurs. Table 6 also notes that the estimates for the control variables indicate that married borrowers with higher education, income, wealth, and credit scores are able to borrow at lower costs.

We then estimate the dependence of the probability of funding success on network centrality with the following logistic regression model:

$$\text{logit}(\text{Funding}_j) = \beta_0 + \beta_1 \text{SocialNetwork}_j + \beta_2 \text{Control}_j + \mu_j + \sigma_t + \varepsilon_j \quad (6)$$

where Funding_j indicates whether loan listing j successfully obtains funding; Network_j is a vector of network centralities for borrower j , including Indegree_j , Outdegree_j , Betweenness_j , Eigenvector_j , and Closeness_j ; Control_j is a set of control variables capturing loans' and borrowers' personal characteristics; and the parameters μ_j and σ_t represent loan and time fixed effects, respectively. In this case, there is no disturbance term, as we assume a generalized linear model with a binomial distribution and a logit link. Table 7 reports the estimation results from the logistic regression.

Columns (1) to (4) in Table 7 show that the coefficients of *Indegree*, *Outdegree*, *Betweenness*, and *Eigenvector* are all positive and significant, signifying that the higher the centrality of borrowers, the higher the chance of funding success. Economically, being at the center of the network, a borrower is able to gain more knowledge and experience than his or her peers, and this helps to set loan conditions that are attractive to investors. However, Column (5) shows that the coefficient of *Closeness* is negative and significant at the 1% level. This indicates that, when a borrower who is close to many borrowers posts a listing, the funding success decreases, possibly due to a competition effect. The control variables of work time, income, and age all show positive and significant effects on the funding probability. However, we find no significant impacts for education, car ownership, and marriage. In terms of loan characteristics, loan requests with a smaller borrowing amount, lower interest rate, and higher credit scores have a higher chance of obtaining funding.

We then examine the effect of networks on loan defaults, using another logistic regression model, as follows:

$$\text{logit}(\text{Default}_j) = \beta_0 + \beta_1 \text{Network}_j + \beta_2 \text{Control}_j + \mu_j + \sigma_t + \varepsilon_j, \quad (7)$$

where Default_j indicates whether loan listing j defaults after successfully obtaining funding. It equals one if the borrower defaults and zero otherwise. Network_j is a vector of network centrality measurements of borrower j . Control_j is a set of control variables, including the borrower's personal characteristics. The parameters μ_j and σ_t represent loan and time fixed effects. Table 8 reports the estimation results.

Table 7: Logistic Regression of Funding Success

	(1) SUCCESS	(2) SUCCESS	(3) SUCCESS	(4) SUCCESS	(5) SUCCESS
Indegree	4.9133*** (11.70)				
Outdegree		0.2821* (1.92)			
Betweenness			7.0916*** (4.40)		
Eigenvector				26.7373*** (12.38)	
Closeness					-17.4545*** (-7.01)
lnAMOUNT	-0.8639*** (-57.91)	-0.8036*** (-57.45)	-0.8029*** (-57.39)	-0.8467*** (-58.23)	-0.8109*** (-57.95)
INTEREST	-0.0982*** (-17.88)	-0.1012*** (-18.26)	-0.0995*** (-17.99)	-0.0982*** (-17.84)	-0.1068*** (-19.00)
MONTHS	0.0026* (1.74)	0.0004 (0.29)	0.0001 (0.10)	0.0023 (1.56)	0.0006 (0.42)
CREDIT	0.1227*** (11.35)	0.1858*** (18.79)	0.1798*** (18.04)	0.1358*** (13.06)	0.2126*** (20.78)
T_length	0.0128*** (8.59)	0.0144*** (9.65)	0.0144*** (9.65)	0.0131*** (8.79)	0.0145*** (9.71)
N_length	-0.0523*** (-13.41)	-0.0541*** (-13.96)	-0.0541*** (-13.96)	-0.0520*** (-13.37)	-0.0561*** (-14.42)
D_length	0.0013*** (10.18)	0.0013*** (10.23)	0.0013*** (10.22)	0.0013*** (10.24)	0.0013*** (10.21)
EDUCATION	-0.0182 (-1.26)	-0.0097 (-0.67)	-0.0100 (-0.70)	-0.0183 (-1.27)	-0.0058 (-0.41)
WORKTIME	0.1481*** (12.00)	0.1483*** (12.06)	0.1512*** (12.30)	0.1495*** (12.13)	0.1426*** (11.60)
INCOME	0.3426*** (29.94)	0.3298*** (29.10)	0.3309*** (29.23)	0.3413*** (29.88)	0.3251*** (28.63)
AGE	0.0188*** (9.41)	0.0215*** (10.78)	0.0215*** (10.77)	0.0194*** (9.70)	0.0219*** (10.97)
HOUSE	-0.1237*** (-5.13)	-0.1015*** (-4.22)	-0.1045*** (-4.35)	-0.1232*** (-5.12)	-0.0933*** (-3.87)
CAR	0.0082 (0.33)	0.0220 (0.88)	0.0197 (0.79)	0.0112 (0.45)	0.0298 (1.19)
Married	-0.0124 (-0.50)	-0.0139 (-0.56)	-0.0150 (-0.60)	-0.0139 (-0.56)	-0.0099 (-0.40)
_cons	7.7978*** (43.58)	7.2014*** (42.41)	7.1754*** (42.30)	7.6096*** (43.59)	7.3534*** (43.02)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
r2_p	0.1587	0.1477	0.1480	0.1552	0.1485

Note: (1) This table reports the logit regression results for funding success. The dependent variable is *SUCCESS*, taking the value one if a loan listing receives full funding and zero otherwise. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *lnAMOUNT*—the natural log of the loan amount (in RMB) that the borrower requests; *MONTHS*—the loan term (in months) that the borrower requests; *CREDIT*—the credit grade of the borrower at the time of creating the listing; *T_Length*—the number of characters in a loan title; *N_Length*—the number of characters in a borrower's nickname; *D_Length*—the number of characters in a loan description; *EDUCATION*—measuring the education level of a borrower; *WORKTIME*—a borrower's work experience measured by years; *INCOME*—the monthly income of a borrower; *AGE*—the age of a borrower expressed in years; *HOUSE*—a dummy variable taking the value one if the borrower is a homeowner and zero otherwise; *CAR*—a dummy variable taking the value one if a borrower owns a car and zero otherwise; *Married*—a dummy variable taking the value one if a borrower is married and zero otherwise; *Year*—a dummy controlling year fixed effect; *Region*—a dummy variable reflecting the area in which a borrower is located; *Industry*—a dummy variable reflecting the industry that a borrower is working in; and *Purpose*—a dummy describing different purposes of borrowing.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the Z-statistics in parentheses. N is the number of observations. r2_p is the pseudo R-square.

Table 8: Logistic Regression of the Default Rate

	(1)	(2)	(3)	(4)	(5)
	DEFAULT	DEFAULT	DEFAULT	DEFAULT	DEFAULT
Indegree	-0.3868 (-0.87)				
Outdegree		-1.7927 (-0.97)			
Betweenness			-12.4933 (-1.16)		
Eigenvector				-7.7392* (-1.90)	
Closeness					-47.3853*** (-3.26)
lnAMOUNT	0.1755*** (4.59)	0.1678*** (4.47)	0.1674*** (4.46)	0.1863*** (4.86)	0.1688*** (4.46)
INTEREST	0.1229*** (9.61)	0.1224*** (9.56)	0.1224*** (9.55)	0.1226*** (9.56)	0.1222*** (9.40)
MONTHS	0.0644*** (23.00)	0.0647*** (23.16)	0.0647*** (23.16)	0.0642*** (22.86)	0.0638*** (22.66)
CREDIT	-2.0658*** (-25.07)	-2.0662*** (-25.07)	-2.0663*** (-25.08)	-2.0650*** (-25.06)	-2.0578*** (-25.06)
T_length	0.0071** (2.37)	0.0069** (2.32)	0.0069** (2.32)	0.0072** (2.41)	0.0074** (2.48)
N_length	0.0082 (1.01)	0.0083 (1.03)	0.0083 (1.03)	0.0082 (1.02)	0.0072 (0.89)
D_length	0.0006** (2.34)	0.0006** (2.34)	0.0006** (2.34)	0.0006** (2.34)	0.0006** (2.26)
EDUCATION	-0.5192*** (-17.18)	-0.5207*** (-17.30)	-0.5206*** (-17.31)	-0.5171*** (-17.11)	-0.5148*** (-17.08)
WORKTIME	-0.0336 (-1.38)	-0.0352 (-1.44)	-0.0353 (-1.44)	-0.0332 (-1.36)	-0.0367 (-1.50)
INCOME	0.2249*** (9.65)	0.2257*** (9.73)	0.2260*** (9.74)	0.2235*** (9.59)	0.2264*** (9.67)
AGE	0.0296*** (7.77)	0.0295*** (7.76)	0.0295*** (7.75)	0.0297*** (7.80)	0.0298*** (7.85)
HOUSE	-0.1529*** (-3.15)	-0.1511*** (-3.11)	-0.1505*** (-3.10)	-0.1526*** (-3.14)	-0.1463*** (-3.00)
CAR	-0.2429*** (-4.68)	-0.2413*** (-4.64)	-0.2407*** (-4.63)	-0.2409*** (-4.64)	-0.2383*** (-4.58)
Married	-0.0495 (-0.99)	-0.0485 (-0.97)	-0.0482 (-0.97)	-0.0508 (-1.02)	-0.0523 (-1.05)
_cons	-3.1682*** (-7.60)	-3.0959*** (-7.49)	-3.0940*** (-7.48)	-3.2556*** (-7.81)	-3.0425*** (-7.26)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	23,365	23,365	23,365	23,365	23,365
r2_p	0.3154	0.3154	0.3155	0.3156	0.3164

Note: (1) This table reports the logit regression results for loan performance. The dependent variable is *DEFAULT*, taking the value one if the funded loan has defaulted and zero otherwise. The explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *lnAMOUNT*—the natural log of the loan amount (in RMB) that the borrower requests; *MONTHS*—the loan term (in months) that the borrower requests; *CREDIT*—the credit grade of the borrower at the time of creating the listing; *T_Length*—the number of characters in a loan title; *N_Length*—the number of characters in a borrower's nickname; *D_Length*—the number of characters in a loan description; *EDUCATION*—measuring the education level of a borrower; *WORKTIME*—a borrower's work experience measured in years; *INCOME*—the monthly income of a borrower; *AGE*—the age of a borrower expressed in years; *HOUSE*—a dummy variable taking the value one if a borrower is a homeowner and zero otherwise; *CAR*—a dummy variable taking the value one if a borrower owns a car and zero otherwise; *Married*—a dummy variable taking the value one if a borrower is married and zero otherwise; *Year*—a dummy controlling the year fixed effect; *Region*—a dummy variable reflecting the area in which a borrower is located; *Industry*—a dummy variable reflecting the industry that a borrower is working in; and *Purpose*—a dummy describing different purposes of borrowing.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the Z-statistics in parentheses. N is the number of observations. r2_p is the pseudo R-square.

Columns (1)–(5) show that the coefficients of *Eigenvector* and *Closeness* are negative and significant, respectively at the 10% and 1% significance levels. The coefficients of the other centrality measures are negative too, although they are not significant. This indicates that borrowers with strong social network links are less likely to default. The effect is stronger for borrowers with higher closeness scores. Those connected to many other borrowers are less likely to default, probably due to the pressure from their peers. In addition, our results indicate that social networks have a certain restrictive effect on borrowers' moral hazard. As expected, borrowers with higher education and wealth (as the ownership of cars and houses reflects) are less likely to default while loans with a larger amount, higher interest rate, longer term, and lower credit score are more likely to default.

To summarize, the empirical evidence presented in this subsection indicates that network centrality improves borrowers' performance by lowering the borrowing costs, enhancing the funding success rate, and moderating the probability of default. In other words, we have empirically verified our *Hypothesis 2*.

4.4 Addressing the Endogeneity Concerns

In evaluating the impact of network centrality on lending behavior and borrowing outcomes, there are a number of important methodological challenges that we need to address. First, some unobservable or omitted variables may contaminate our estimation results. For example, investor sentiment or behavior bias may change the borrowing outcome (Zhang and Liu 2012; Lin and Viswanathan 2016). We employ the 2SLS and the probit instrumental variable (IV) model to address this concern. Second, as default depends on success, we can only observe the defaults among the borrowers who were successfully in their loan requests and cannot observe defaults among those who failed to raise the requested funds. Hence, our estimation of the defaults might be susceptible to sample selection bias. To moderate this bias, we employ the Heckman selection model (Heckman 1979). Third, individual effects that are not observable over time might also affect a lender's lending behavior. For example, lenders can learn from past lending experience. In this regard, we introduce a positive exogenous impact on the industry and use a difference-in-difference approach to investigate the effect of network centrality on lending behavior.

4.4.1 Instrumental Variable Estimation

The first robustness challenge of this study is that our estimation might be susceptible to the bias arising from unobservable variables. For example, investor sentiment or behavior bias, although unobservable, are likely to exert an impact on borrowing outcomes (Zhang and Liu 2012; Lin and Viswanathan 2016). To address this concern, one potential solution is to find an instrumental variable (IV) that is correlated with the network centrality of the borrowers but does not directly affect the interest rate and the probability of funding success.

Such instrumental variables can be derived from peer effects. The literature has well recognized the important role of peers in forming financial decisions. For example, Leary and Roberts (2014) acknowledged that firms' financing decisions are in large part responses to the financing decisions of peer firms. We borrow from these studies and develop instruments named *Me_Indegree*, *Me_Outdegree*, *Me_Betweenness*, *Me_Eigenvector*, and *Me_Closeness* for model identification. They correspond to the average network centrality of borrowers with a similar educational level and length of work experience. We believe that the network centrality of peers with similar characteristics will affect the network centrality of an individual borrower but not the

borrower’s performance. Table 9 shows how the estimation results change after introducing the instrumental variables.

Table 9: 2SLS and IV Probit Estimation Results

	(1)	(2)	(3)	(4)	(5)
	Indegree	Outdegree	Betweenness	Eigenvector	Closeness
Me_Indegree	0.81*** (10.51)				
Me_Outdegree		0.81*** (12.88)			
Me_Betweenness			0.75*** (16.10)		
Me_Eigenvector				0.80*** (12.47)	
Me_Closeness					0.65*** (15.62)
Indegree					
Outdegree					
Betweenness					
Eigenvector					
Closeness					
lnAMOUNT	0.01*** (15.91)	-0.00*** (-3.75)	-0.00*** (-2.76)	0.00*** (14.72)	-0.00*** (-11.96)
INTEREST	-0.00*** (-13.00)	-0.01*** (-17.94)	-0.00*** (-18.17)	-0.00*** (-14.68)	-0.00*** (-18.51)
MONTHS	-0.00 (-0.39)	0.00*** (11.78)	0.00*** (11.93)	-0.00** (-2.13)	0.00** (2.19)
CREDIT	0.03*** (29.30)	0.02*** (20.08)	0.00*** (21.37)	0.00*** (30.55)	0.00*** (44.16)
T_length	0.00*** (7.22)	-0.00 (-1.11)	-0.00 (-0.45)	0.00*** (6.53)	0.00 (1.52)
N_length	-0.00*** (-11.41)	-0.00*** (-9.82)	-0.00*** (-5.71)	-0.00*** (-11.84)	-0.00*** (-14.59)
D_length	0.00*** (2.90)	0.00*** (3.85)	0.00*** (3.49)	0.00*** (3.10)	0.00 (0.48)
EDUCATION	-0.00*** (-6.83)	-0.00*** (-8.21)	-0.00*** (-10.70)	-0.00*** (-8.08)	-0.00*** (-6.94)
WORKTIME	-0.01*** (-12.46)	-0.00*** (-10.49)	-0.00*** (-9.65)	-0.00*** (-12.41)	-0.00*** (-13.93)
INCOME	-0.01*** (-11.45)	-0.01*** (-14.09)	-0.00*** (-13.41)	-0.00*** (-12.64)	-0.00*** (-12.26)
AGE	0.00*** (9.36)	0.00*** (3.54)	0.00** (2.40)	0.00*** (9.22)	0.00*** (3.95)

continued on next page

Table 9 *continued*

	(1)	(2)	(3)	(4)	(5)
	Indegree	Outdegree	Betweenness	Eigenvector	Closeness
HOUSE	0.01*** (11.67)	0.01*** (10.83)	0.00*** (10.61)	0.00*** (12.82)	0.00*** (9.83)
CAR	0.01*** (8.87)	0.01*** (11.00)	0.00*** (8.22)	0.00*** (9.71)	0.00*** (8.35)
Married	0.00** (2.19)	0.01*** (7.89)	0.00*** (4.27)	0.00*** (5.90)	0.00*** (6.14)
_cons	-0.03*** (-3.05)	0.11*** (13.70)	0.01*** (12.87)	-0.00 (-1.31)	0.01*** (18.61)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
F	106.19	17.98	21.89	93.23	197.31
r2_a	0.31	0.19	0.21	0.30	0.44
	(6)	(7)	(8)	(9)	(10)
	INTEREST	INTEREST	INTEREST	INTEREST	INTEREST
Me_Indegree					
Me_Outdegree					
Me_Betweenness					
Me_Eigenvector					
Me_Closeness					
Indegree	-1.49** (-2.42)				
Outdegree		-4.57*** (-4.61)			
Betweenness			-59.80*** (-5.86)		
Eigenvector				-14.16*** (-2.80)	
Closeness					-60.75*** (-2.60)
lnAMOUNT	0.16*** (12.02)	0.13*** (11.05)	0.13*** (10.71)	0.16*** (12.23)	0.12*** (8.22)
INTEREST					
MONTHS	0.07*** (51.95)	0.08*** (53.37)	0.08*** (52.46)	0.07*** (50.89)	0.07*** (51.00)
CREDIT	-0.29*** (-15.13)	-0.24*** (-11.91)	-0.19*** (-7.22)	-0.27*** (-12.60)	-0.23*** (-5.57)
T_length	0.02*** (14.43)	0.02*** (13.40)	0.02*** (13.12)	0.02*** (14.54)	0.02*** (14.28)
N_length	-0.02*** (-4.38)	-0.02*** (-5.28)	-0.02*** (-5.08)	-0.02*** (-4.56)	-0.02*** (-4.62)

continued on next page

Table 9 *continued*

	(6)	(7)	(8)	(9)	(10)
	INTEREST	INTEREST	INTEREST	INTEREST	INTEREST
D_length	0.00 (1.28)	0.00** (2.12)	0.00** (2.38)	0.00 (1.41)	0.00 (1.08)
EDUCATION	-0.14*** (-9.39)	-0.13*** (-8.72)	-0.13*** (-8.66)	-0.14*** (-9.31)	-0.14*** (-8.91)
WORKTIME	0.03** (2.10)	-0.01 (-0.82)	-0.04** (-2.21)	0.02* (1.77)	0.01 (0.80)
INCOME	-0.04*** (-3.70)	-0.06*** (-5.01)	-0.07*** (-5.66)	-0.05*** (-3.97)	-0.05*** (-4.02)
AGE	0.00 (1.53)	0.00 (1.53)	0.00* (1.76)	0.00* (1.66)	0.00 (1.53)
HOUSE	-0.11*** (-4.42)	-0.09*** (-3.50)	-0.07** (-2.42)	-0.11*** (-4.12)	-0.10*** (-3.73)
CAR	-0.18*** (-7.07)	-0.14*** (-5.18)	-0.13*** (-4.57)	-0.18*** (-6.69)	-0.17*** (-6.02)
Married	-0.15*** (-5.72)	-0.12*** (-4.67)	-0.12*** (-4.69)	-0.14*** (-5.40)	-0.13*** (-4.94)
_cons	9.68*** (64.23)	10.04*** (65.66)	10.07*** (65.92)	9.67*** (64.64)	10.13*** (52.34)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
F					
r2_a	0.26	0.27	0.25	0.27	0.27
	(11)	(12)	(13)	(14)	(15)
	SUCCESS	SUCCESS	SUCCESS	SUCCESS	SUCCESS
Me_Indegree					
Me_Outdegree					
Me_Betweenness					
Me_Eigenvector					
Me_Closeness					
Indegree	1.34*** (2.84)				
Outdegree		1.34** (2.18)			
Betweenness			2.19 (0.34)		
Eigenvector				8.82*** (2.58)	
Closeness					6.05 (0.41)
InAMOUNT	-0.49*** (-48.28)	-0.47*** (-52.61)	-0.47*** (-59.66)	-0.49*** (-50.50)	-0.47*** (-43.05)

continued on next page

Table 9 continued

	(11)	(12)	(13)	(14)	(15)
	SUCCESS	SUCCESS	SUCCESS	SUCCESS	SUCCESS
INTEREST	−0.06*** (−16.31)	−0.05*** (−9.70)	−0.06*** (−10.93)	−0.06*** (−15.34)	−0.06*** (−10.00)
MONTHS	0.00 (0.12)	−0.00 (−1.46)	−0.00 (−0.70)	0.00 (0.08)	−0.00 (−0.79)
CREDIT	0.10*** (7.93)	0.09*** (7.29)	0.11*** (7.45)	0.10*** (7.37)	0.10*** (4.08)
T_length	0.01*** (9.01)	0.01*** (9.99)	0.01*** (9.85)	0.01*** (9.04)	0.01*** (9.78)
N_length	−0.03*** (−13.17)	−0.03*** (−11.33)	−0.03*** (−13.35)	−0.03*** (−12.86)	−0.03*** (−10.50)
D_length	0.00*** (10.15)	0.00*** (9.14)	0.00*** (9.90)	0.00*** (10.13)	0.00*** (10.04)
EDUCATION	−0.01 (−1.13)	−0.01 (−1.47)	−0.01 (−0.95)	−0.01 (−1.21)	−0.01 (−1.07)
WORKTIME	0.09*** (11.03)	0.10*** (10.85)	0.09*** (8.28)	0.09*** (10.96)	0.09*** (10.33)
INCOME	0.19*** (26.04)	0.20*** (27.69)	0.19*** (23.87)	0.19*** (25.22)	0.19*** (24.98)
AGE	0.01*** (10.05)	0.01*** (10.43)	0.01*** (10.78)	0.01*** (10.02)	0.01*** (10.42)
HOUSE	−0.07*** (−4.32)	−0.07*** (−4.70)	−0.06*** (−3.93)	−0.07*** (−4.37)	−0.06*** (−4.08)
CAR	0.01 (0.59)	−0.00 (−0.11)	0.01 (0.67)	0.01 (0.61)	0.01 (0.50)
Married	−0.01 (−0.60)	−0.02 (−1.07)	−0.01 (−0.68)	−0.01 (−0.58)	−0.01 (−0.79)
_cons	4.41*** (40.67)	4.05*** (29.60)	4.21*** (35.88)	4.37*** (42.78)	4.15*** (23.56)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
F					
r2_a					

Note: (1) This table reports the 2SLS regression results for the interest rate and the IV probit regression results for the probability of funding success. The dependent variables are: (i) *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); (ii) INTEREST, the interest rate that the borrower pays on the loan; and (iii) *SUCCESS*, taking the value one if a loan listing receives full funding and zero otherwise. The explanatory variables include: *Me_Indegree*, *Me_Outdegree*, *Me_Betweenness*, *Me_Eigenvector*, and *Me_Closeness*, which is the network centrality of a borrower's peer; *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *InAMOUNT*—the natural log of the loan amount (in RMB) that the borrower requested; *MONTHS*—the loan term (in months) that the borrower requests; *CREDIT*—the credit grade of the borrower at the time of creating the listing; *T_Length*—the number of characters in a loan title; *N_Length*—the number of characters in a borrower's nickname; *D_Length*—the number of characters in a loan description; *EDUCATION*—measuring the education level of a borrower; *WORKTIME*—a borrower's work experience measured in years; *INCOME*—the monthly income of a borrower; *AGE*—the age of a borrower expressed in years; *HOUSE*—a dummy variable taking the value one if a borrower is a homeowner and zero otherwise; *CAR*—a dummy variable taking the value one if a borrower owns a car and zero otherwise; *Married*—a dummy variable taking the value one if a borrower is married and zero otherwise; *Year*—a dummy controlling the year fixed effect; *Region*—a dummy variable reflecting the area in which a borrower is located; *Industry*—a dummy variable reflecting the industry that a borrower is working in; and *Purpose*—a dummy describing the different purposes of borrowing.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report the T/Z-statistics in parentheses. N is the number of observations. F is the F statistic. r2_a is the adjusted R-square

The first-stage regression results, which Columns (1)–(5) of Table 9 present, show a positively significant coefficient for the instrumental variable (*Me_Indegree*, *Me_Outdegree*, *Me_Betweenness*, *Me_Eigenvector*, and *Me_Closeness*), meaning that the higher the network centrality of the peers, the higher the network centrality of the borrower. Moreover, the F-statistic that appears at the bottom is equal to 106.193, 17.981, 21.891, 93.228, and 197.313, respectively. According to Staiger and Stock (1994), the suggested critical F-value is 8.96 when the number of instruments is one. With the F-statistic being much greater than 10, we can reject the null hypotheses that the coefficients for the instruments are significantly different from zero at the 1% level, thus excluding the concern about weak instrumental variables. The second-stage regression results, which Columns (6) to (15) present, show that the network centrality variables still have a significantly negative effect on the interest rate and a significantly positive effect on the probability of funding success. These results are in line with the estimation without the instrumental variables, confirming the robustness of our conclusion.

4.4.2 Heckman Selection Model

The second methodological challenge of this study is that we can only observe the defaults of the borrowers who have successfully obtained the funds that they requested and not the defaults of those who failed to raise the funds. Our estimation results for the default response might thus be contaminated by a sample selection bias. We can employ Heckman's (1979) selection model to address this concern.

We can achieve a valid implementation of the Heckman selection model by introducing an exogenous variable (or instrument) that we can include in the first-stage regression model with funding success as a response variable but exclude from the second-stage regression with default as a response variable (see e.g. Little 1985). We believe that the higher the number of lenders bidding for a loan, the more likely it is that it will successfully obtain funding. On the other hand, the number of bidders should not directly affect the actual default rate of a loan. We can thus employ the total number of bidders for a loan (*BIDS*) as an instrument for model identification. Table 10 shows how the estimation results change after introducing the instrumental variable.

Table 10: Heckman Two-Step Regression on the Probability of Default

	(1)	(2)	(3)	(4)	(5)
	SUCCESS	SUCCESS	SUCCESS	SUCCESS	SUCCESS
BIDS	0.083*** (30.56)	0.084*** (30.66)	0.083*** (30.65)	0.083*** (30.60)	0.084*** (30.62)
Indegree	0.874*** (7.69)				
Outdegree		0.634*** (6.65)			
Betweenness			6.656*** (6.03)		
Eigenvector				6.029*** (7.43)	
Closeness					3.510** (2.29)

continued on next page

Table 10 *continued*

	(1)	(2)	(3)	(4)	(5)
	SUCCESS	SUCCESS	SUCCESS	SUCCESS	SUCCESS
IMR_Indegree					
IMR_Outdegree					
IMR_Betweenness					
IMR_Eigenvector					
IMR_Closeness					
lnAMOUNT	-0.727*** (-66.16)	-0.718*** (-65.89)	-0.719*** (-65.89)	-0.726*** (-66.02)	-0.717*** (-65.84)
INTEREST	-0.026*** (-8.17)	-0.025*** (-7.90)	-0.025*** (-7.74)	-0.026*** (-8.13)	-0.027*** (-8.55)
MONTHS	0.006*** (6.22)	0.006*** (5.83)	0.006*** (5.70)	0.006*** (6.29)	0.006*** (6.11)
CREDIT	0.045*** (6.54)	0.054*** (8.25)	0.051*** (7.69)	0.045*** (6.59)	0.058*** (8.14)
T_length	0.009*** (7.99)	0.009*** (8.59)	0.009*** (8.58)	0.009*** (8.07)	0.009*** (8.46)
N_length	-0.025*** (-9.19)	-0.026*** (-9.65)	-0.027*** (-9.76)	-0.025*** (-9.25)	-0.027*** (-9.71)
D_length	0.000** (1.98)	0.000** (2.04)	0.000** (2.02)	0.000** (1.97)	0.000** (2.14)
EDUCATION	-0.000 (-0.02)	0.004 (0.40)	0.005 (0.48)	0.001 (0.09)	0.005 (0.52)
WORKTIME	0.055*** (6.66)	0.056*** (6.85)	0.058*** (7.02)	0.055*** (6.68)	0.052*** (6.32)
INCOME	0.090*** (10.68)	0.086*** (10.24)	0.086*** (10.31)	0.090*** (10.72)	0.084*** (9.99)
AGE	0.004** (2.53)	0.004*** (2.67)	0.004*** (2.69)	0.004** (2.52)	0.004*** (2.71)
HOUSE	-0.036** (-2.13)	-0.029* (-1.75)	-0.032* (-1.89)	-0.035** (-2.11)	-0.027 (-1.61)
CAR	-0.018 (-1.07)	-0.017 (-0.98)	-0.018 (-1.04)	-0.018 (-1.07)	-0.013 (-0.75)
Married	0.015 (0.90)	0.016 (0.96)	0.015 (0.93)	0.014 (0.87)	0.017 (1.06)
_cons	6.220*** (50.10)	6.139*** (49.62)	6.144*** (49.63)	6.210*** (50.06)	6.158*** (49.66)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	44,481	44,481	44,481	44,481	44,481
r2_p	0.503	0.502	0.502	0.503	0.501

continued on next page

Table 10 *continued*

	(6)	(7)	(8)	(9)	(10)
	DEFAULT	DEFAULT	DEFAULT	DEFAULT	DEFAULT
BIDS					
Indegree	-0.289 (-0.66)				
Outdegree		-1.813 (-0.95)			
Betweenness			-12.511 (-1.13)		
Eigenvector				-7.032* (-1.74)	
Closeness					-46.433*** (-3.23)
IMR_Indegree	0.131 (1.64)				
IMR_Outdegree		0.131* (1.65)			
IMR_Betweenness			0.131* (1.65)		
IMR_Eigenvector				0.122 (1.52)	
IMR_Closeness					0.118 (1.48)
lnAMOUNT	0.176*** (4.59)	0.170*** (4.52)	0.170*** (4.51)	0.187*** (4.86)	0.171*** (4.50)
INTEREST	0.119*** (9.12)	0.118*** (9.08)	0.118*** (9.07)	0.119*** (9.09)	0.118*** (8.93)
MONTHS	0.065*** (23.03)	0.065*** (23.18)	0.065*** (23.18)	0.064*** (22.88)	0.064*** (22.67)
CREDIT	-2.063*** (-25.02)	-2.063*** (-25.02)	-2.063*** (-25.03)	-2.063*** (-25.01)	-2.055*** (-25.00)
T_length	0.007** (2.50)	0.007** (2.46)	0.007** (2.46)	0.008** (2.53)	0.008*** (2.61)
N_length	0.007 (0.81)	0.007 (0.83)	0.007 (0.82)	0.007 (0.83)	0.006 (0.70)
D_length	0.001** (2.41)	0.001** (2.41)	0.001** (2.41)	0.001** (2.40)	0.001** (2.32)
EDUCATION	-0.521*** (-17.22)	-0.522*** (-17.34)	-0.522*** (-17.33)	-0.518*** (-17.15)	-0.516*** (-17.10)
WORKTIME	-0.030 (-1.22)	-0.031 (-1.27)	-0.031 (-1.27)	-0.030 (-1.21)	-0.033 (-1.36)
INCOME	0.231*** (9.77)	0.231*** (9.84)	0.232*** (9.85)	0.229*** (9.69)	0.231*** (9.76)
AGE	0.030*** (7.84)	0.030*** (7.84)	0.030*** (7.83)	0.030*** (7.87)	0.030*** (7.92)

continued on next page

Table 10 *continued*

	(6)	(7)	(8)	(9)	(10)
	DEFAULT	DEFAULT	DEFAULT	DEFAULT	DEFAULT
HOUSE	-0.157*** (-3.23)	-0.155*** (-3.20)	-0.155*** (-3.19)	-0.157*** (-3.22)	-0.150*** (-3.08)
CAR	-0.244*** (-4.71)	-0.243*** (-4.67)	-0.242*** (-4.66)	-0.242*** (-4.67)	-0.239*** (-4.60)
Married	-0.048 (-0.96)	-0.047 (-0.95)	-0.047 (-0.94)	-0.049 (-0.99)	-0.051 (-1.02)
_cons	-3.206*** (-7.66)	-3.151*** (-7.58)	-3.149*** (-7.58)	-3.292*** (-7.87)	-3.091*** (-7.34)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	23,365	23,365	23,365	23,365	23,365
r2_p	0.316	0.316	0.316	0.316	0.317

Note: (1) This table reports the Heckman two-step regression result for the probability of default. In columns (1)–(5), the dependent variable is *SUCCESS*, taking the value 1 if a loan listing obtains full funding and zero otherwise. In columns (6)–(10), the dependent variable is the *DEFAULT* dummy, taking the value one if the funded loan has defaulted and zero otherwise. *IMR_Indegree*, *IMR_Outdegree*, *IMR_Betweenness*, *IMR_Eigenvector*, and *IMR_Closeness* are the inverse Mills ratio. *BIDS* is the number of lenders in a loan. Other explanatory variables include: *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender’s network centrality (see section 3.2.1 for the definition); *lnAMOUNT*—the natural log of the loan amount (in RMB) that the borrower requests; *INTEREST*—the interest rate that the borrower pays on the loan; *MONTHS*—the loan term (in months) that the borrower requests; *CREDIT*—the credit grade of the borrower at the time of creating the listing; *T_Length*—the number of characters in a loan title; *N_Length*—the number of characters in a borrower’s nickname; *D_Length*—the number of characters in a loan description; *EDUCATION*—measuring the education level of a borrower; *WORKTIME*—a borrower’s work experience measured in years; *INCOME*—the monthly income of a borrower; *AGE*—the age of a borrower expressed in years; *HOUSE*—a dummy variable taking the value one if the borrower is a homeowner and zero otherwise; *CAR*—a dummy variable taking the value one if a borrower owns a car and zero otherwise; *Married*—a dummy variable taking the value one if a borrower is married and zero otherwise; *Year*—a dummy controlling the year fixed effect; *Region*—a dummy variable reflecting the area in which a borrower is located; *Industry*—a dummy variable reflecting the industry that a borrower is working in; and *Purpose*—a dummy describing different purposes of borrowing.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and present the T/Z-statistics in parentheses. N is the number of observations. r2_p is the pseudo R-square.

Columns (1)–(5) of Table 10 report the first-step estimate of *SUCCESS*. The coefficients for *BIDS* are positive and significant, implying that the larger the number of lenders, the higher the likelihood of a borrower obtaining funding following a loan application. Columns (6)–(10) present the second-stage estimate for the defaults in which we add the Mill’s ratio centrality measures, that is, the fitted values of the centrality measures from the first-stage regression (*IMR_Indegree*, *IMR_Outdegree*, *IMR_Betweenness*, *IMR_Eigenvector*, and *IMR_Closeness*). The coefficients of *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness* are significant and similar in size to the baseline results in Table 8, meaning that our conclusions are robust after controlling for sample selection bias.

4.4.3 Quasi-natural Experiments

We further evaluate the robustness of our model by introducing a policy announced and implemented by the PRC government as an exogenous shock. On 18 July 2015, the People's Bank of China, together with ten ministries, jointly issued *The Guiding Opinions on Promoting the Healthy Development of Internet Finance*. This was the first time that the PRC government had shown a positive signal to support the development of P2P lending. Before that, many people were concerned that the industry could be banned if the government did not recognize it. We treat the announcement as an exogenous and positive shock to the industry development, because it encouraged people to participate in P2P lending. We then examine the impact of that shock on the relationship between a lender's network centrality and his or her lending behavior. Intuitively, we expect that this positive and exogenous shock will strengthen the information advantage of lenders and encourage them to be more active in making investments.

To verify this expectation, we perform a difference-in-difference analysis (Nunn and Qian 2011) by estimating the following equation:

$$\begin{aligned} Lending_i = & \beta_0 + \beta_1 Network_i \times POST_t + \beta_2 Network_i + \beta_3 POST_t \\ & + \beta_4 Control_i + \varepsilon_i, \end{aligned} \quad (8)$$

where the dependent variable measures the lending behavior of lender i with three different indicators of $MONEY_i$, $LENDTIME_i$, and $FSTB_M_i$, respectively. $Network_i$ is a network centrality vector for lender i , which contains five variables: $Indegree_i$, $Outegree_i$, $Betweenness_i$, $Eigenvector_i$, and $Closeness_i$. $Control_i$ is a set of control variables reflecting lenders' personal characteristics. Finally, ε_i is a random disturbance term. The new component of the model is $POST_t$, a dummy variable that equals one for the post-policy period and zero for the pre-policy period. We expect the coefficient for the interaction term $Network_i \times POST_t$ to be consistent with the results of the baseline regressions. Table 11 shows the estimation results.

Columns (1)–(5) in Table 11 report the impacts of network centrality on $MONEY$; Columns (6)–(10) report the impacts of network centrality on $LENDTIME$; and Columns (11)–(15) report the impacts of network centrality on $FSTB_M$. The coefficients of all the interaction items between network centrality measurements and $POST$ are consistent with the baseline results in terms of both the significance level and the sign, further validating our finding that network centrality encourages lenders to invest in P2P lending in larger amounts and at a faster pace.

Table 11: Difference-in-Differences Test Results

	(1)	(2)	(3)	(4)	(5)
	MONEY	MONEY	MONEY	MONEY	MONEY
Indegree_POSIT	-304.06*** (-4.77)				
Indegree	79.50*** (4.25)				
Outdegree_POSIT		-301.23*** (-4.28)			
Outdegree		174.16*** (8.39)			
Betweenness_POSIT			-2,481.51*** (-3.35)		
Betweenness			865.37*** (4.91)		
Eigenvector_POSIT				-4,402.04*** (-5.15)	
Eigenvector				2,056.51*** (9.65)	
Closeness_POSIT					479.83 (0.64)
Closeness					-3,757.01*** (-13.02)
POST	616.45 (1.48)	619.02 (1.49)	616.93 (1.48)	615.61 (1.48)	626.04 (1.50)
L_EDUCATION	32.76*** (7.01)	26.96*** (5.74)	32.57*** (6.97)	29.45*** (6.29)	42.80*** (9.14)
L_WORKTIME	-1.91 (-0.48)	1.75 (0.44)	-1.29 (-0.33)	-1.58 (-0.40)	-2.94 (-0.75)
L_INCOME	70.35*** (19.52)	68.56*** (19.02)	71.16*** (19.84)	70.25*** (19.57)	69.67*** (19.46)
L_AGE	13.76*** (19.05)	13.69*** (18.95)	13.76*** (19.05)	13.86*** (19.19)	13.82*** (19.15)
L_HOUSE	39.20*** (4.00)	43.39*** (4.44)	41.20*** (4.21)	35.65*** (3.64)	46.22*** (4.72)
L_CAR	73.92*** (8.81)	68.81*** (8.13)	73.45*** (8.73)	67.55*** (8.03)	96.06*** (11.28)
L_Married	-93.97*** (-8.95)	-103.12*** (-9.92)	-97.99*** (-9.44)	-91.89*** (-8.82)	-105.00*** (-10.10)
_cons	-208.96* (-1.79)	-15.45 (-0.13)	-22.79 (-0.20)	-10.24 (-0.09)	-33.57 (-0.29)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
r2_a	0.36	0.36	0.36	0.36	0.36

continued on next page

Table 11 *continued*

	(6)	(7)	(8)	(9)	(10)
	LENDTIME	LENDTIME	LENDTIME	LENDTIME	LENDTIME
Indegree_POSIT	0.02*** (12.04)				
Indegree	-0.02*** (-13.82)				
Outdegree_POSIT		0.01*** (4.45)			
Outdegree		-0.01*** (-6.33)			
Betweenness_POSIT			0.15*** (7.40)		
Betweenness			-0.16*** (-10.01)		
Eigenvector_POSIT				0.10*** (3.91)	
Eigenvector				-0.11*** (-5.44)	
Closeness_POSIT					0.02 (0.96)
Closeness					-0.01 (-0.70)
POST	0.02* (1.68)	0.02 (1.52)	0.02 (1.62)	0.02 (1.53)	0.02 (1.48)
L_EDUCATION	-0.00*** (-5.05)	-0.00*** (-4.00)	-0.00*** (-5.05)	-0.00*** (-4.86)	-0.00*** (-5.81)
L_WORKTIME	0.00*** (7.35)	0.00*** (5.72)	0.00*** (6.59)	0.00*** (6.66)	0.00*** (6.53)
L_INCOME	-0.00*** (-14.65)	-0.00*** (-15.82)	-0.00*** (-16.17)	-0.00*** (-16.47)	-0.00*** (-17.10)
L_AGE	-0.00*** (-12.57)	-0.00*** (-11.95)	-0.00*** (-12.34)	-0.00*** (-12.28)	-0.00*** (-11.93)
L_HOUSE	-0.00*** (-4.35)	-0.00*** (-7.78)	-0.00*** (-6.22)	-0.00*** (-7.04)	-0.00*** (-8.10)
L_CAR	-0.00 (-0.34)	-0.00 (-1.50)	-0.00 (-0.82)	-0.00 (-1.45)	-0.00*** (-2.60)
L_Married	0.00*** (7.74)	0.01*** (13.20)	0.01*** (10.73)	0.01*** (12.19)	0.01*** (13.60)
_cons	0.03*** (4.39)	0.09*** (17.46)	0.09*** (17.01)	0.09*** (17.43)	0.09*** (17.82)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
r2_a	0.37	0.37	0.37	0.37	0.37

continued on next page

Table 11 *continued*

	(11)	(12)	(13)	(14)	(15)
	FSTB_M	FSTB_M	FSTB_M	FSTB_M	FSTB_M
Indegree_POSIT	557.34*** (26.79)				
Indegree	-555.19*** (-27.75)				
Outdegree_POSIT		260.35*** (16.10)			
Outdegree		-281.85*** (-18.44)			
Betweenness_POSIT			4,462.01*** (21.86)		
Betweenness			-4,670.45*** (-24.95)		
Eigenvector_POSIT				3,260.71*** (17.20)	
Eigenvector				-3,415.85*** (-19.81)	
Closeness_POSIT					2,123.28*** (10.63)
Closeness					-2,162.48*** (-11.45)
POST	215.68 (1.30)	155.54 (0.96)	194.26 (1.17)	164.88 (0.99)	134.44 (0.81)
L_EDUCATION	4.91** (2.33)	10.72*** (4.98)	5.08** (2.39)	6.81*** (3.19)	5.25** (2.52)
L_WORKTIME	14.36*** (7.18)	4.98** (2.45)	10.23*** (5.06)	10.63*** (5.26)	9.27*** (4.56)
L_INCOME	-20.51*** (-11.69)	-26.16*** (-14.33)	-27.47*** (-15.13)	-28.98*** (-15.74)	-32.65*** (-17.27)
L_AGE	-2.31*** (-7.00)	-1.80*** (-5.37)	-2.14*** (-6.46)	-2.10*** (-6.30)	-1.76*** (-5.25)
L_HOUSE	-17.12*** (-3.16)	-67.96*** (-11.58)	-42.97*** (-7.76)	-54.30*** (-9.55)	-74.07*** (-12.41)
L_CAR	-13.47*** (-3.02)	-27.39*** (-6.01)	-18.14*** (-4.03)	-24.60*** (-5.40)	-32.97*** (-7.13)
L_Married	42.53*** (7.64)	131.40*** (18.95)	87.95*** (14.40)	112.10*** (17.25)	138.82*** (19.62)
_cons	1,069.41*** (8.51)	2,482.79*** (29.06)	2,415.43*** (28.81)	2,473.69*** (29.10)	2,532.27*** (29.45)
Loan Fixed Effects	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Day of Week Fixed Effects	YES	YES	YES	YES	YES
Hour of Day Fixed Effects	YES	YES	YES	YES	YES
N	273,508	273,508	273,508	273,508	273,508
r2_a	0.97	0.97	0.97	0.97	0.97

Note: (1) This table reports the OLS regression results of the investment amount. The dependent variable is (i) *MONEY*, the amount that a lender invests in each investment. (ii) *LENDTIME* is the percentage of the amount that the listing requests when lenders invest. (iii) *FSTB_M* is the time interval between the posting of the loan listing and the lender's investment. The explanatory variables include *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*—the lender's network centrality (see section 3.2.1 for the definition); *POST* is a dummy that equals one for the post-event period and zero for the pre-event period; *L_EDUCATION* measures the education level of a lender; *L_WORKTIME* is a lender's work experience measured in years; *L_INCOME*—the monthly income of a lender; *L_AGE*—the age of a lender expressed in years; *L_HOUSE*—a dummy variable taking the value one if a lender is a homeowner and zero otherwise; *L_CAR*—a dummy variable taking the value one if a lender owns a car and zero otherwise; and *L_Married*—a dummy variable taking the value one if a lender is married and zero otherwise; all the regressions include loan fixed effects, year fixed effects, day-of-week fixed effects, and hour-of-day fixed effects.

(2) *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. We use robust standard errors and report T-statistics in parentheses. N is the number of observations. r2_a is the adjusted R-square.

5. CONCLUSIONS

The importance of social networks in the financial market motivated us to measure the centrality of lenders and borrowers on the P2P lending platform and explore their roles in shaping the market equilibrium.

Employing data from Renrendai, a leading lending platform in the PRC, we first gauge the position of each lender and borrower in the network by means of five centrality measures, specifically *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *Closeness*. Our empirical analysis reveals that the lenders who are more central in the network not only invest larger amounts but also invest more swiftly than their peers, reflecting the information advantage arising from their position in the network. Furthermore, the borrowers who are more central in the network are able to borrow at lower interest rates and with higher success rates. At the same time, they are less likely to default. These findings imply that network linkages not only spur the flow of information among peers but also improve market efficiency and encourage reputation protection in the P2P lending market. These data-driven results echo the existing literature that has underlined the importance of a social network in shaping investment decisions and facilitating financial transactions. Our findings also suggest that it is possible to use centrality measures to infer the creditworthiness of borrowers. In the evolving online credit markets, information asymmetry remains serious and might even increase. As reliable credit scores are not available for most people, P2P lending platforms should consider integrating network measurements into their credit evaluation systems.

Our research has important policy implications for financial regulators in their attempt to protect the consumers of digital financial services and maintain financial stability. While financial technologies effectively improve the convenience and accessibility of financial services, they also trigger new risks to the existing financial architecture. Therefore, regulators and supervisors need to keep the regulatory technology up to date. Our research suggests that social network analysis, combined with explainable regression models, can effectively advance our understanding of the effect of network centrality on borrowers' outcomes and on lenders' investment behavior. Furthermore, it is possible to apply network centrality to forecast the probability of default, which is critical for risk monitoring and prevention.

REFERENCES

- Ahern, K.R. 2017. "Information Networks: Evidence from Illegal Insider Trading Tips." *Journal of Financial Economics* 125: 26–47.
- Ambrus, A., M. Mobius, and A. Szeidl. 2014. "Consumption Risk-Sharing in Social Networks." *American Economic Review* 104: 149–82.
- Bajo, E., T.J. Chemmanur, K. Simonyan, and H. Tehranian. 2016. "Underwriter Networks, Investor Attention, and Initial Public Offerings." *Journal of Financial Economics* 122: 376–408.
- Barabási, A.-L. 2016. *Network Science*. Cambridge: Cambridge University Press.
- Bonacich, P. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2: 113–20.
- Borgatti, S.P., A. Mehra, D.J. Brass, and G. Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323: 892–5.
- Bourles, R., Y. Bramouille, and E. Perez-Richet. 2017. "Altruism in Networks." *Econometrica* 85: 675–89.
- Breza, E., and A.G. Chandrasekhar. 2019. "Social Networks, Reputation, and Commitment: Evidence from a Savings Monitors Experiment." *Econometrica* 87: 175–216.
- Brown, J.R., Z. Ivkovic, P.A. Smith, and S. Weisbenner. 2008. "Neighbors Matter: Causal Community Effects and Stock Market Participation." *Journal of Finance* 63: 1509–31.
- Butler, A.W., and U.G. Gurun. 2012. "Educational Networks, Mutual Fund Voting Patterns, and CEO Compensation." *Review of Financial Studies* 25: 2533–62.
- Cai, J., and A. Szeidl. 2018. "Interfirm Relationships and Business Performance." *Quarterly Journal of Economics* 133: 1229–82.
- Cai, Y., and M. Sevilir. 2012. "Board Connections and M&A Transactions." *Journal of Financial Economics* 103: 327–49.
- Calomiris, C.W., and M. Carlson. 2017. "Interbank Networks in the National Banking Era: Their Purpose and Their Role in the Panic of 1893." *Journal of Financial Economics* 125: 434–53.
- Chen, X., B. Huang, and D. Ye. 2018. "The Role of Punctuation in P2P Lending: Evidence from China." *Economic Modelling* 68: 634–43.
- Cohen, L., A. Frazzini, and C. Malloy. 2008. "The Small World of Investing: Board Connections and Mutual Fund Returns." *Journal of Political Economy* 116: 951–79.
- Cohen, L., A. Frazzini, and C. Malloy. 2010. "Sell-Side School Ties." *Journal of Finance* 65: 1409–37.
- Cruz, C., J. Labonne, and P. Querubin. 2017. "Politician Family Networks and Electoral Outcomes: Evidence from the Philippines." *American Economic Review* 107: 3006–37.
- Duarte, J., S. Siegel, and L. Young. 2012. "Trust and Credit: The Role of Appearance in Peer-to-Peer Lending." *Review of Financial Studies* 25: 2455–83.

- Duchin, R., and D. Sosyura. 2013. "Divisional Managers and Internal Capital Markets." *Journal of Finance* 68: 387–429.
- El-Khatib, R., K. Fogel, and T. Jandik. 2015. "CEO Network Centrality and Merger Performance." *Journal of Financial Economics* 116: 349–82.
- Engelberg, J., P.J. Gao, and C.A. Parsons. 2012. "Friends with Money." *Journal of Financial Economics* 103: 169–88.
- . 2013. "The Price of a CEO's Rolodex." *Review of Financial Studies* 26: 79–114.
- Faley, O., T. Kovacs, and A. Venkateswaran. 2014. "Do Better-Connected CEOs Innovate More?" *Journal of Financial and Quantitative Analysis* 49: 1201–25.
- Fang, L.H., and S. Huang. 2017. "Gender and Connections among Wall Street Analysts." *Review of Financial Studies* 30: 3305–35.
- Feigenberg, B., E. Field, and R. Pande. 2013. "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance." *Review of Economic Studies* 80: 1459–83.
- Ferreira, M.A., and P. Matos. 2012. "Universal Banks and Corporate Control: Evidence from the Global Syndicated Loan Market." *Review of Financial Studies* 25: 2703–44.
- Fracassi, C. 2017. "Corporate Finance Policies and Social Networks." *Management Science* 63: 2420–38.
- Fracassi, C., and G. Tate. 2012. "External Networking and Internal Firm Governance." *Journal of Finance* 67: 153–94.
- Freeman, L.C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1: 215–39.
- Fuster, A., M. Plosser, P. Schnabl, and J. Vickery. 2019. "The Role of Technology in Mortgage Lending." *Review of Financial Studies* 32: 1854–99.
- Garmaise, M.J., and T.J. Moskowitz. 2003. "Informal Financial Networks: Theory and Evidence." *Review of Financial Studies* 16: 1007–40.
- Granovetter, M.S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78: 1360–80.
- Grullon, G., S. Underwood, and J.P. Weston. 2014. "Comovement and Investment Banking Networks." *Journal of Financial Economics* 113: 73–89.
- Gurun, U.G., N. Stoffman, and S.E. Yonker. 2018. "Trust Busting: The Effect of Fraud on Investor Behavior." *Review of Financial Studies* 31: 1341–76.
- Han, B., and L.Y. Yang. 2013. "Social Networks, Information Acquisition, and Asset Prices." *Management Science* 59: 1444–57.
- Hasan, I., C.K. Hoi, Q. Wu, and H. Zhang. 2017. "Social Capital and Debt Contracting: Evidence from Bank Loans and Public Bonds." *Journal of Financial and Quantitative Analysis* 52: 1017–47.
- Haselmann, R., D. Schoenherr, and V. Vig. 2018. "Rent Seeking in Elite Networks." *Journal of Political Economy* 126: 1638–90.
- Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153–61.

- Hildebrand, T., M. Puri, and J. Rocholld. 2017. "Adverse Incentives in Crowdfunding." *Management Science* 63: 587–608.
- Hochberg, Y.V., A. Ljungqvist, and Y. Lu. 2007. "Whom You Know Matters: Venture Capital Networks and Investment Performance." *Journal of Finance* 62: 251–301.
- . 2010. "Networking as a Barrier to Entry and the Competitive Supply of Venture Capital." *Journal of Finance* 65: 829–59.
- Hong, H., J.D. Kubik, and J.C. Stein. 2005. "Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers." *Journal of Finance* 60: 2801–24.
- Hong, H., and J.M. Xu. 2019. "Inferring Latent Social Networks from Stock Holdings." *Journal of Financial Economics* 131: 323–44.
- Houston, J.F., J. Lee, and F. Suntheim. 2018. "Social Networks in the Global Banking Sector." *Journal of Accounting & Economics* 65: 237–69.
- Ishii, J., and Y.H. Xuan. 2014. "Acquirer–Target Social Ties and Merger Outcomes." *Journal of Financial Economics* 112: 344–63.
- Iyer, R., A.I. Khwaja, E.F.P. Luttmer, and K. Shue. 2016. "Screening Peers Softly: Inferring the Quality of Small Borrowers." *Management Science* 62: 1554–77.
- Iyer, R., and M. Puri. 2012. "Understanding Bank Runs: The Importance of Depositor–Bank Relationships and Networks." *American Economic Review* 102: 1414–45.
- Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl. 2009. "Trust and Social Collateral." *Quarterly Journal of Economics* 124: 1307–61.
- Khanna, V., E.H. Kim, and Y. Lu. 2015. "CEO Connectedness and Corporate Fraud." *Journal of Finance* 70: 1203–52.
- Kilduff, M., C. Crossland, W. Tsai, and M.T. Bowers. 2016. "Magnification and Correction of the Acolyte Effect: Initial Benefits and Ex Post Settling up in NFL Coaching Careers." *Academy of Management Journal* 59: 352–75.
- Kuhnen, C.M. 2009. "Business Networks, Corporate Governance, and Contracting in the Mutual Fund Industry." *Journal of Finance* 64: 2185–220.
- Lauritzen, S.L. 1996. *Graphical Models*. Oxford: Oxford University Press.
- Leary, M.T., and M.R. Roberts. 2014. "Do Peer Firms Affect Corporate Financial Policy?" *Journal of Finance* 69: 139–78.
- Lin, M.F., N.R. Prabhala, and S. Viswanathan. 2013. "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending." *Management Science* 59: 17–35.
- Lin, M.F., and S. Viswanathan. 2016. "Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market." *Management Science* 62: 1393–414.
- Little, R.J. 1985. "A Note about Models for Selectivity Bias." *Econometrica* 53: 1469–74.
- Liu, D., D.J. Brass, Y. Lu, and D.Y. Chen. 2015. "Friendships in Online Peer-to-Peer Lending: Pipes, Prisms, and Relational Herding." *MIS Quarterly* 39: 729–42.
- Michels, J. 2012. "Do Unverifiable Disclosures Matter? Evidence from Peer-to-Peer Lending." *Accounting Review* 87: 1385–413.

- Moretti, E., 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." *Review of Economic Studies* 78: 356–93.
- Munshi, K., and M. Rosenzweig. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural–Urban Wage Gap." *American Economic Review* 106: 46–98.
- Nguyen, B.D. 2012. "Does the Rolodex Matter? Corporate Elite's Small World and the Effectiveness of Boards of Directors." *Management Science* 58: 236–52.
- Nunn, N., and N. Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." *Quarterly Journal of Economics* 126: 593–650.
- Ozsoylev, H.N., J. Walden, M.D. Yavuz, and R. Bildik. 2014. "Investor Networks in the Stock Market." *Review of Financial Studies* 27: 1323–66.
- Paravisini, D., V. Rappoport, and E. Ravina. 2017. "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios." *Management Science* 63: 279–97.
- Pons, P., and M. Lathapy. 2006. "Computing Communities in Large Networks Using Random Walks." *Journal of Graph Algorithms and Applications* 10 (2): 191–218.
- Pool, V.K., N. Stoffman, and S.E. Yonker. 2015. "The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolios." *Journal of Finance* 70: 2679–732.
- Pope, D.G., and J.R. Sydnor. 2011. "What's in a Picture? Evidence of Discrimination from Prosper.com." *Journal of Human Resources* 46: 53–92.
- Rossi, A.G., D. Blake, A. Timmermann, I. Tonks, and R. Wermers. 2018. "Network Centrality and Delegated Investment Performance." *Journal of Financial Economics* 128: 183–206.
- Schmidt, B. 2015. "Costs and Benefits of Friendly Boards during Mergers and Acquisitions." *Journal of Financial Economics* 117: 424–47.
- Schoenherr, D. 2019. "Political Connections and Allocative Distortions." *Journal of Finance* 74: 543–86.
- Shue, K. 2013. "Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers." *Review of Financial Studies* 26: 1401–42.
- Sorenson, O., V. Assenova, G.C. Li, J. Boada, and L. Fleming. 2016. "Expand Innovation Finance via Crowdfunding." *Science* 354: 1526–8.
- Staiger, D., and J.H. Stock. 1994. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65: 557–86.
- Stanton, R., J. Walden, and N. Wallace. 2018. "Mortgage Loan Flow Networks and Financial Norms." *Review of Financial Studies* 31: 3595–642.
- Strausz, R. 2017. "A Theory of Crowdfunding: A Mechanism Design Approach with Demand Uncertainty and Moral Hazard." *American Economic Review* 107: 1430–76.
- Stuart, T.E., and S. Yim. 2010. "Board Interlocks and the Propensity to Be Targeted in Private Equity Transactions." *Journal of Financial Economics* 97: 174–89.
- Tang, H. 2019. "Peer-to-Peer Lenders Versus Banks: Substitutes or Complements?" *Review of Financial Studies* 32: 1900–38.

- Vallée, B., and Y. Zeng. 2019. "Marketplace Lending: A New Banking Paradigm?" *Review of Financial Studies* 32: 1939–82.
- Wei, Z.Y., and M.F. Lin. 2017. "Market Mechanisms in Online Peer-to-Peer Lending." *Management Science* 63: 4236–57.
- Zhang, J.J., and P. Liu. 2012. "Rational Herding in Microloan Markets." *Management Science* 58: 892–912.