Identifying Features for Detecting Fraudulent Loan Requests on P2P Platforms

Jennifer Xu
Department of Computer Information Systems
Bentley University
Waltham, MA, U.S.A.
jxu@bentley.edu

Dongyu Chen
School of Business
Soochow University
Suzhou, Jiangsu, China
chendongyu@suda.edu.cn

Michael Chau
School of Business
The University of Hong Kong
Hong Kong, China
mchau@business.hku.hk

***Abstract*—This exploratory study is intended to address the problem of fraudulent loan requests on peer-to-peer (P2P) platforms. We propose a set of features that capture the behavioral characteristics (e.g., learning, past performance, social networking, and herding manipulation) of malevolent borrowers, who intentionally create loan requests to acquire funds from lenders but default later on. We found that using the widely adopted classification methods such as Random Forest and Support Vector Machines, the proposed feature set outperform the baseline feature set in helping detect fraudulent loan requests. Although the performance (e.g., *Recall* or *Sensitivity*) is still not up to its optimum, this study demonstrates that by analyzing the transaction records of confirmed malevolent borrowers, it is possible to capture some useful behavioral patterns for fraud detection. Such features and methods would possibly help lenders identify loan request frauds and avoid financial losses.**

Keywords—P2P lending, fraud detection, feature selection

I. Introduction

The past ten years have witnessed a rapid growth of the peer-to-peering (P2P) lending platforms all over the world. In the United States, a total of $22 billion loans have been originated on the two largest P2P platforms, Prosper and Lending Club [1, 2]. In China, there have emerged more than 4,000 P2P platforms by the end of 2015 and originated around ¥2,470 billion loans [3]. Typically hosted on a website, a P2P platform provides a virtual marketplace in which individuals can acquire loans directly from other people without depending on an intermediate financial institution such as a bank. This new type of financial venue may benefit borrowers who have difficulty obtaining funds from banks. In the meanwhile, it makes it possible for lenders to achieve a higher return than investing on traditional financial products such as bank CDs and municipal bonds. As a result, P2P lending has attracted millions of users across the countries since its inception and has been regarded as one of the most important financial innovations in the past decade.

However, a higher return is naturally associated with a higher risk. Specifically, lenders may suffer from loan default, which happens when a borrower fails to repay the lenders. Indeed, a borrower may even intentionally create a fraudulent loan request to cheat the lenders out of their money in the first place. For example, an enterpriser may submit a loan request to fund his/her business project that does not actually exist. An extremist may pretend to seek help with his/her financial difficulty while the true purpose is to prepare for a terrorist attack. In 2014, four paper companies in Guangzhou, China, committed fraudulent loans, causing a P2P platform, My089, a total of ¥100 million bad debt risk [4]. Right before the 2015 San Bernardino attack, with a loan request for debt consolidation, the terrorist acquired $28,000 from Proposer.com to purchase weapons and explosives [5]. To enhance the chance of funding success, a borrower may employ several tactics to manipulate the lenders' decision-making so as to attract them to invest on his/her loans. After receiving the full amount and the loan is materialized, the borrower would make no payments or only partial payments to the lenders during the repayment period, causing the lenders financial losses.

To reduce lenders' financial risks and ensure the orderly functioning of P2P marketplaces, it is important to develop measures and methods to help detect possible fraudulent loan requests. However, such research potentially faces several challenges. First, because a platform can only get and provide limited information about borrowers, it is difficult to find a rich set of features for signaling fraudulent loan requests. Second, the training of fraud detection models and techniques relies on the knowledge of the ground truth, which often is not available. Third, as P2P lending is a relatively new economic phenomenon, there has not been much research that offers useful directions and frameworks for P2P fraud detection.

To bridge the research gaps, this exploratory study aims at proposing features that may help identify possible fraudulent loan requests during the loan auction process. The proposed features focus on the behavioral aspects of borrowers and seek to capture the distinguishing characteristics of malevolent borrowers who are most likely to commit malicious default.

The remainder of this paper is organized as follows. The next section reviews the literature on P2P lending and online fraud detection. Section 3 introduces our proposed features. Section 4 presents the data and methods. The classification results are reported and discussed in Section 5. The last section concludes the paper and plans for future work.

Xu, J. J., Chen, D., and Chau, M. "Identifying Features for Detecting Fraudulent Loan Requests on P2P Platforms," in Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI 2016), Tucson, Arizona, USA, Sep 27-30, 2016.

II. RELATED WORK

A. P2P Lending

The risk of fraudulent loan requests is rooted in the very nature of peer-to-peer transactions. With the absence of financial intermediaries the impact of information asymmetry [6] is substantially elevated. In traditional lending transactions, the lender is not an individual but a financial institution. With the access to comprehensive records of the borrower's credit history and with the help of advanced risk assessment instruments and tools, an institutional lender can often accurately predict the borrower's repayment capability and default risks. In the P2P context, however, individual lenders do not have complete and accurate information about the identity and credit history of a borrower. In this marketplace, where only limited information (e.g., credit scores and debt-income ratios) about borrowers is available, the problem of information asymmetry makes lenders very vulnerable to loan request fraud. Moreover, in countries such as China, which does not yet have a credit system, a borrower who commits fraud on a platform may receive no more penalty than being unable to borrow money only on the particular platform in the future.

Consequently, a rational lender usually strikes to seek information about a prospective borrower to reduce risks resulted from information asymmetry and uncertainty. According to the financial information theory [7, 8], two types of information (hard and soft) exist in financial markets. Hard information refers to financial characteristics that can be easily quantified (e.g., credit scores and debt-to-income ratio); while soft information refers to the information that is difficult to capture and summarize using a numeric value (e.g., trustworthiness, social relationship).

In the P2P lending context, lenders have largely relied on hard information to decide to whom to lend their money. They generally tend to invest on loans requested by borrowers with good financial records and credit histories [9, 10]. For instance, Khwaja et al. found that lenders' decisions are significantly affected by borrowers' default rate, debt-to-income ratio, and the number of loan requests in the past six months [11]. It has also been found that lenders prefer culturally similar and geographically proximate borrowers [12], and that this pattern of home bias is common in different situations based on the analysis of detailed transaction data and a natural experiment on Prosper.com [13].

Although hard to collect and process, soft information can also play an important role in lenders' investment decision-making. For instance, Lin et al. discovered that a borrower's social network has a significant effect on lenders' decisions [10]. Duarte et al. found that borrowers who appear to be more trustworthy in their pictures are more likely to get their loan requests funded [9]. Similarly, Pope and Sydnor studied several individual characteristics (e.g., race, gender, age, attractiveness, etc.) based on borrower images and found that black borrowers are less likely to receive funding than white borrowers with similar credit profiles [14].

In addition, a lender's decision may be influenced by other lenders' behavior, a phenomenon known as herding. Herzenstein et al. studied the strategic herding behavior by lenders and concluded that lenders are more likely to bid on a loan that already has received bids [15]. Similarly, Zhang and Liu discovered the evidence of rational herding among lenders, especially in the early stage (e.g., the first day) in the bidding process [16]. Such herding behavior has also been found on P2P platforms in other countries such as Korea and China [17, 18].

Factors that affect funding outcomes of loan requests have been used to analyze and predict loan performance, an important measure of which is default rate. Most studies have reported that borrowers with better credit grades tend to have lower default rates. Lin et al. estimated that the odds of a borrower's defaulting decreases by 9% on average if the borrower has friends in their social network with verified identities and who act as lenders [10]. However, other studies have found that borrowers with good personal or social capital may not necessarily perform well in their ex-post loan payments and tend to default more often than those with good hard scores [19, 20].

B. Online Fraud Detection

Fraud has raised serious concerns in the financial industry and academia. Fraudulent transactions are seen not only in offline markets but also in online markets such as e-commerce sites. The unique characteristics of the Internet, such as low entry barriers, user anonymity, and spatial and temporal separation between users, have made it a fertile field for deception and fraud [21]. Prior research has focused on identifying features and developing techniques for online fraud detection.

In the e-commerce context, product rating histories have been used to help detect fraudulent or fake product ratings and reviews. Cai and Zhu proposed to examine the deviation of individual ratings from the majority or the past ratings of a product [22]. However, if the majority of the early raters are fraudulent, the results will be misleading. An alternative approach is to identify regular raters. Teacy et al. utilized regular raters' ratings to filter out suspicious raters by examining the dissimilarity [23]. This method also has its flaw: if fraudulent raters strategically behave like the regular raters, this method may become invalid. Some other scholars proposed to examine the received rating series of each product and filter out the product under fraud attacks. A clustering based method is then applied to discriminate fraudulent raters.

In the P2P lending literature, only a few studies are related to the issue of fraud detection. For instance, when studying a crowdfunding projects, Agrawal et al. discovered that a large number of investors are project initiators' friends and family members, who play a significant role in forging the investors' geography effect in the market [24]. Gao and Lin examined the text descriptions of loan requests and identified several linguistic cues for detecting deceptive loan requests and predicting loan default [25]. Such vocal and linguistic cues have been used to detect financial fraud in offline markets [26]. Xu et al. proposed to use big data approach and leverage multiple data sources to detect P2P loan request fraud. However, no empirical results have been reported [27].

In terms of the fraud detection techniques, supervised learning methods are often employed to detect fraud in various contexts (e.g., credit card fraud, insurance fraud, and medical fraud). Commonly used methods include Decision Trees, Support Vector Machines, Genetic Algorithms, Bayesian Belief Networks, and Neural Networks. The three most widely adopted performance metrics for financial fraud detection are accuracy, sensitivity, and specificity [28].

III. Proposed Features

We propose a set of features for helping detect possible fraudulent loan requests (also called listings) on P2P platforms. These features are intended to capture the characteristics of an individual borrower's behavior in his/her transaction history, social network, and loan auction (also called bidding) process. We focus on funded listings by "blacklisted" individuals, who have been confirmed to be deceitful borrowers and defaulted maliciously on the platform under study. By deceitfully pledging to repay yet defaulting later, these borrowers have committed fraud by definition [29].

We investigate these borrowers' behavior in four aspects: learning, past performance, social networking, and herding manipulation.

A. Learning

It may not be easy for a first-time borrower to successfully get his/her listing funded on a platform. From the perspective of a borrower, who decides to borrow and then default, the ultimate goal is to appropriate as much money as possible. Therefore, he/she must first learn how to create a listing and experiment with different borrowing amounts and interest rates in order to maximize the likelihood of funding success. Thus, we use three features to capture a borrower's learning behavior:

- *#_Prev_Listings*: the total number of listings created by the borrower before the current listing;
- *Amt_Prev_Listings*: the total amount requested by the borrower before the current listing;
- *#_Funded_Listings*: the total number of listings, created by the borrower, that have been successfully funded before the current listing.

B. Past Performance

Generally, to attract bids from lenders, borrowers need to convince the lenders that they are honest and trustworthy. To engage lenders' trust, a borrower may purposefully maintain a good track record by repaying previous loans on time. Because the platform under study has implemented its own credit assessment system based on historical transaction records on the platform, it is not uncommon for borrowers to request small-amount loans simply for "building records" or "increasing credit grade". A blacklisted borrower may have adopted this practice in order to demonstrate his/her trustworthiness and repayment capability. We propose a feature to reflect this aspect of a borrower's past performance by the time a new listing is created:

- *%_Repaid*: the percentage of fully repaid loans out of all funded listings by the borrower before the current listing.

On the other hand, although a borrower might have fully repaid previous loans, there may have been a trace of delinquency (i.e., default, late payment, partial payment) in his/her past monthly payment records:

- *#_Delinquency*: the number of delinquent monthly payments by the borrower before the current listing.

C. Social Networking

It has become well known that an individual's social network can to a certain extent affect his/her chance of getting funded. The endorsement made by a borrower's friends may serve as a quality signal and enhance the funding success [10]. On the platform under study, an individual can connect with another one and specify the type of the relationship (e.g., friend, family, colleague, schoolmate, acquaintance, or online friend):

- *#_Friends*: the number of friends (of all types) of the borrower before the current listing.

Furthermore, a blacklisted borrower may purposefully request or even hire his/her friends to bid on his/her listing to increase the number of endorsements and supports. However, because each bid must be at least ¥50 on the platform under study, the borrower may not be able to request many friends, especially acquaintance and online friends, to cooperate. To minimize the cost, the borrower may only afford to hire close friends and family members to bid on the listing. Therefore, we pay special attention to the bids made by a borrower's family and close friends (FF) during the bidding process:

- *#_Bids_FF*: the number of bids made by family members, relatives, and friends of the borrower on the current listing;
- *Amt_FF*: the total amount invested by family members, relatives, and friends of the borrower on the current listing.

D. Herding Manipulation

As reported in the P2P literature, herding behavior exists on P2P platforms and lenders are more likely to invest on listings that have already received bids from others [16]. From a borrower's perspective, if he/she can manage to manipulate the bidding process by creating a herding momentum on the first day, the chance of funding success may be substantially increased. Although difficult, it is not completely impossible for a borrower to secretly hire some people, like an Internet water army, to make fake bids on his/her listing. We propose two features to measure the herding momentum in the early stage of a bidding process:

- *#_Bids_1stDay*: the number of bids received by the listing on the first day;
- *Amt_1stDay*: the total amount received by the listing on the first day.

IV. Data and Methods

A. Data

The platform under study, MyLending,[1] is one of the largest P2P lending marketplaces in China. Launched in 2007, MyLending has attracted over 17 million users and more than ¥100 million loans have been funded. Because China does not have a nationwide credit system, MyLending has implemented its own credit assessment system by evaluating each borrower's background information (e.g., education level and degrees, professional certificates, etc.) and previous repayment records on MyLending, if applicable, and assigning each borrower a letter credit grade from A (High Quality) to HR (High Risk).

MyLending provided a proprietary sample that consists of all 39,694 listings made by 23,049 borrowers on the platform from June 2007 to December 2011. The sample also contains 582,201 records (e.g., lender, bidding date and time, bidding amount, etc.) about all bids made on listings during this period. Among these listings, 9,771 (24%) were fully funded and the remaining 29,923 (76%) listings failed to receive sufficient funds. The payment information includes only monthly payments that were due or made by borrowers between August 7, 2010 and August 25, 2011. Thus, the resulting sample in this study contains only the 6,562 successfully funded and materialized loans, of which the payment information is available during this 13-month period.

B. Sample Labeling

Our greatest challenge when labeling the sample was the lack of knowledge about the ground truth. Note that although a borrower might have defaulted during a loan's repayment period, which usually ranges between 3 and 12 months, we cannot just conclude that this particular listing was a fraudulent one when it was created. A borrower may fail to make a monthly payment on time because of various reasons (e.g., unemployment, medical emergency, etc.) but make a lump sum payment (with late penalties) eventually. Therefore, we could not simply label all loans with delinquent payments as fraudulent.

We addressed this problem by consulting the "Blacklist" published by MyLending, which exposes malevolent borrowers, who refused to repay their loans even after the platform had exhausted all payment collection measures at its disposal (e.g., email reminders, phone calls, or third-party debt collectors). Each entry on this list contains a borrower's user ID, real name, default amount, and the year in which the borrower defaulted. Note that if a borrower later repays a loan in full, his/her entry will be removed from the list immediately. We found that 103 borrowers, who defaulted in 2010 or 2011 according to this list, had 691 materialized loans in our sample (691/6562 = 10.5%). Among the 691 listings, 270 loans were fully repaid. Eventually we labeled 217 listings out of the remaining 421 loans (51.5%) as fraudulent.[2]

C. Methods and Metrics

We intended to find out if the proposed features could offer additional clues for signaling fraudulent loan requests besides the baseline features: the borrower's credit grade, the listing's requested amount, the interest offered, repayment period length, and the loan category (i.e., the purpose of the loan). These baseline features have typically been used in prior P2P lending studies.

We employed two supervised learning methods, Random Forest (RF) and Support Vector Machines (SVM), to classify fraudulent listings. Both methods have been shown to have outstanding classification performance [30]. A widely adopted data mining software tool, Weka [31], was used to train and test the classifiers.

As in most classification studies, we selected *Accuracy* to measure the performance of a classifier. However, because only 217 listings (3.3% of the 6,562 loans) are labeled as fraud, the accuracy can be as high as 96.7% even without using any feature (i.e., all listings are blindly classified as normal, nonfraudulent ones). As a result, we also selected *Precision, Recall*, and *F-Measure*, and the corresponding metrics for the fraud class to measure how well a feature set sorts out fraudulent listings. In addition, we selected to report the area under the ROC curve to compare the discriminating capabilities of the two feature sets.

$Precision_{Fraud}$ = # Correctly identified fraudulent listings/# Listings classified as fraudulent,

$Recall_{Fraud}$ = # Correctly identified fraudulent listings/# True fraudulent listing,

$F\text{-}Measure_{Fraud} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$.

Note that $Precision_{Fraud}$ is essentially 1 minus *False Positive Rate*; and $Recall_{Fraud}$ is *Sensitivity*.

V. Results and Discussion

A. Blacklisted vs. Regular Borrowers

As the assumption underlying these proposed features is that blacklisted borrowers may exhibit different behavioral characteristics than regular borrowers, we first compared the means of these features between the two groups of borrowers. Table 1 reports the means and standard deviations (in parentheses) of these features. It can be seen that except for three features (*%_repaid*, *#_bids_FF*, *Amt_FF*), blacklisted and regular borrowers have significantly different behaviors measured by these features (p < 0.001). For example, on average, while regular borrowers only have about 5.6 listings before creating a new listing, blacklisted borrowers have created more than 13 listings, demonstrating a strong learning and trial behavior. The distributions of this feature are also quite different between the two groups of borrowers (see Figure 1). While most regular borrowers have only a few previous listings, a few regular borrowers have a large number of listings. This power-law distribution is rather common in

[1] To ensure confidentiality, we use a fictitious company name in this study.

[2] These 217 loans' payments had long been overdue by the end of the 13-month period. We also compared the total overdue amount of each borrower of these loans with his/her total default amount on the blacklist and further verified these loans' status.

most large systems [32]. However, the distribution for the blacklisted borrowers deviates drastically from a typical linear power-law line on the log-log plot.

TABLE I. MEANS OF THE PROPOSED FEATURES FOR THE BLACKLISTED AND REGULAR BORROWERS

	Blacklisted Borrowers	Regular Borrowers
#_Prev_Listings	**13.4*** (8.4)	5.6 (5.9)
Amt_Prev_Listings	**42,808*** (50,522)	20,654 (40,213)
#_Funded_Listings	**9.9*** (6.3)	4.0 (4.3)
%_Repaid	51 (34.4)	53 (39.6)
#_Delinquency	**3.7*** (3.4)	0.89 (2.21)
#_Friends	**43.6*** (56.7)	15.3 (38.7)
#_Bids_FF	0.52 (1.53)	0.28 (1.36)
Amt_FF	237 (887)	136 (1,244)
#_Bids_1stday	**18.6*** (10.1)	13.2 (13.6)
Amt_1stday	**4,811*** (3,736)	3,271 (8,636)

* $p < 0.001$

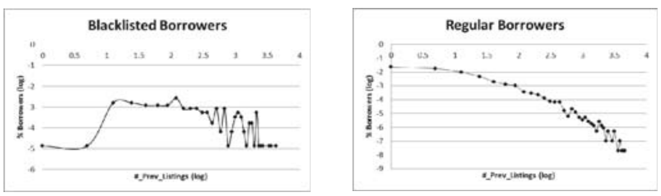

Fig. 1. The log-log plots of the distributions of *#_Prev_Listings* for blacklisted (left) and regular (right) borrowers

Loan requests by blacklisted borrowers tended to be slightly more likely (77%) to get fully funded than those by the regular borrowers (71%). Their percentages of previously fully repaid loans were not significantly different (both around 50%). However, the blacklisted borrowers did have significantly more traces of delinquency (3.7) than the regular borrowers (0.89) did.

It is interesting that although blacklisted borrowers had significantly more friends (43.6) than regular borrowers (15.3), they did not necessarily receive more bids and investment from their families and friends. In other words, the two features (*#_bids_FF* and *Amt_FF*) would fail to capture the conspiracy, if there was one, between a borrower and his/her family members and friends. One possible reason is that these blacklisted borrowers might not necessarily have identified their family members and friends on MyLending.

The two first-bidding-day features for capturing herding momentum appear to be significantly different between the two groups of borrowers with the blacklisted borrowers receiving significantly more bids (18.8) and higher amounts ($4,811) than the regular borrowers (13.2 and $3,271, respectively). Thus, it was possible (with no confirmed evidence) that some of the first-day bids made on the listings created by the blacklisted borrowers might have been fake, aiming to lure other lenders to invest on their loans.

B. *Detecting Fraudulent Listings*

We used RF and SVM to classify the sample using the baseline feature set and the proposed feature set, which includes both the proposed and baseline features. Because the sample was extremely unbalanced with only 3.3% of the listings labeled as fraud, any classification method assuming a balanced sample would perform poorly. Thus, we employed the cost sensitive procedure in Weka to increase the penalty on a failure to identify fraudulent cases. By adjusting the cost parameter in the procedure, we observed that the *F-Measure* reached its optimum when the cost was set to be 4.0 using RF and 10.0 using SVM.

We then set the cost parameters to their optimal values in the two classifiers respectively. Table 2 reports the metric values between the proposed feature set and the baseline feature set. These metric values are means out of 10-fold cross validation. Table 2 shows that the proposed feature set performs significantly better than the baseline feature set in *Accuracy* and *F-Measure* ($p < 0.001$) using both classification methods. Between the two methods, RF shows slightly better performance in detecting fraudulent listings.

TABLE II. PERFORMANCE COMPARISON BETWEEN THE PROPOSED FEATURES AND THE BASELINE FEATURES

		RF	SVM
Accuracy	Proposed	**0.96***	0.94
	Baseline	0.94	**0.97***
Precision	Proposed	0.96	0.95
	Baseline	0.95	0.94
Recall	Proposed	0.96	0.94
	Baseline	0.95	**0.97***
F-Measure	Proposed	0.96	0.94
	Baseline	0.95	0.95
Precision $_{Fraud}$	Proposed	**0.41***	**0.21**
	Baseline	0.14	N/A
Recall $_{Fraud}$	Proposed	**0.33***	**0.33***
	Baseline	0.05	0.0
F-Measure $_{Fraud}$	Proposed	**0.36***	**0.26**
	Baseline	0.07	N/A
Area under ROC	Proposed	**0.66***	**0.64***
	Baseline	0.56	0.50

* $p < 0.001$

It can be seen that, with the RF method, while only approximately 5% of fraudulent listings were correctly identified using the baseline features, about one third of the listings were correctly identified using the proposed feature set (see *Recall*$_{Fraud}$). When with the SVM, the baseline set completely missed all the fraudulent listings. That is, the SVM classified all listings as regular ones using the baseline features. Therefore, although its overall *Accuracy*, *Recall*, and *F-Measure* are higher than the proposed set, it is of no use for detecting fraud.

The overall *Precision*, *Recall*, and *F-Measure* values remain to be high with RF. The area under ROC curve for the proposed feature set is greater than that for the baseline feature set, which is barely better than a random prediction. However, with respect to the fraud class, the *Precision*, *Recall*, and *F-Measure* (< 0.5) are not yet satisfactory. Nonetheless, although the proposed feature set still missed two thirds of the fraudulent listings, they did provide more information, which can be extracted from the data available to the lenders, for detecting loan request frauds on the P2P platform.

VI. Conclusions

This study aims at exploring the problem of fraudulent loan requests on P2P platforms. We propose a set of features that capture the behavioral characteristics (e.g., learning, past performance, social networking, and herding manipulation) of malevolent borrowers, who create listings to cheat lenders out of their money. We found that using the widely adopted classification methods such as Random Forest and SVM, the proposed feature set outperform the baseline features in detecting frauds. Although the performance (e.g., *Recall* or *Sensitivity*) is still not satisfactory, this exploratory study demonstrates that by analyzing the behavior of confirmed malevolent borrowers, it is possible to capture some useful patterns for fraud detection. Such instruments and tools would possibly help lenders keep vigilant to loan request fraud and avoid financial losses. Future studies may explore and discover more features that reflect borrower behavioral characteristics and creditability.

References

[1] Lending Club. (2016, May 16, 2016). *LendingClub.com*. Available: www.lendingclub.com

[2] Proposer. (2016, May 16, 2016). *Prosper.com*. Available: www.prosper.com

[3] X. Liu. (2016, May 16, 2016). *Annual Report on P2P Websites*. Available: http://www.qianzhan.com/qzdata/detail/149/160426-676209e7.html

[4] Sina Finance. (2014, May 16, 2016). *P2P Bad Debt Record at My089*. Available: http://finance.sina.com.cn/chanjing/gsnews/20140829/015720153605.shtml

[5] J. R. Koren and J. Puzzanghera, "Loan to San Bernardino shooter draws scrutiny to online lending industry," in *Los Angeles Times*, ed. Los Angeles, CA, 2015.

[6] G. A. Akerlof, "The market for 'lemons': Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics,* vol. 84, pp. 488-500, 1970.

[7] M. A. Atkinson and C. Kydd, "Individual characteristics associated with World Wide Web use: An empirical study of playfulness and motivation," *ACM SIGMIS Database,* vol. 28, pp. 53-62, 1997.

[8] M. A. Pertersen, *Information: Hard and Soft*. Chicago, IL: Northwestern University, 2002.

[9] J. Duarte*, et al.*, "Trust and credit: The role of appearance in peer-to-peer lending," *The Review of Financial Studies,* vol. 25, pp. 2455-2484, 2012.

[10] M. Lin*, et al.*, "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," *Management Science,* vol. 59, pp. 17-35, 2013.

[11] A. I. Khwaja*, et al.*, "Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?," John F. Kennedy School of Government, Harvard University2009.

[12] G. Burtch*, et al.*, "Cultural differences and geography as determinants of online prosocial lending," *MIS Quarterly,* vol. 38, pp. 773-794, 2014.

[13] M. Lin and S. Viswanathan, "Home bias in online investments: An empirical study of an online crowd funding market," *Management Science,* Forthcoming.

[14] D. G. Pope and J. R. Sydnor, "What's in a picture? Evidence of discrimination from Prosper.com," *The Journal of Human Resources,* vol. 46, pp. 53-92, 2011.

[15] M. Herzenstein*, et al.*, "Strategic herding behavior in peer-to-peer loan auctions," *Journal of Interactive Marketing,* vol. 25, pp. 27-36, 2011.

[16] J. Zhang and P. Liu, "Rational herding in microloan markets," *Management Science,* vol. 58, pp. 892-912, 2012.

[17] D. Chen and Z. Lin, "Rational or irrational herding in online microloan markets: Evidence from China," *Working Paper,* Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425047, 2015.

[18] E. Lee and B. Lee, "Herding behavior in online P2P lending: An empirical investigation," *Electronic Commerce Research and Applications,* vol. 11, pp. 495-503, 2012.

[19] S. Berger and F. Gleisner, "Emergence of financial intermediaries in electronic markets: The case of online P2P lending," *Business Research,* vol. 2, pp. 39-65, 2009.

[20] S. Herrero-Lopez, "Social Interactions in P2P Lending," in *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, 2009.

[21] B. Xiao and I. Benbasat, "Product-related deception in e-commerce: A theoretical perspective," *MIS Quarterly,* vol. 35, pp. 169-196, 2011.

[22] Y. Cai and D. Zhu, "Rating fraud detection: Towards designing a trustworthy reputation systems," in *Thirty Sixth International Conference on Information Systems*, 2015.

[23] W. Teacy*, et al.*, "Travos: trust and reputation in the context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems,* vol. 12, pp. 183-198, 2006.

[24] A. K. Agrawal*, et al.* (2011). *The Geography of Crowfunding*.

[25] Q. Gao and M. F. Lin, "Lemon or cherry: The value of texts in debt crowdfunding," *Working Paper,* Available: http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2446114, 2015.

[26] C. S. Throckmorton*, et al.*, "Financial fraud detection using vocal, linguistic and financial cues," *Decision Support Systems,* vol. 74, pp. 78-87, 6// 2015.

[27] J. Xu*, et al.*, "P2P lending fraud detection: A big data approach," in *Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics*, Ho Chi Minh City, Vietnam, 2015, pp. 71-81.

[28] J. West and M. Bhattacharya, "Intelligent financial fraud detection: a comprehensive review," *Computers & Security,* vol. 12, 2015.

[29] Merriam-Webster, *Merriam-Webster's Dictionary of Law*. Springfield, MA: Merriam-Webster, Inc., 2016.

[30] J. Han*, et al.*, *Data Mining: Concepts and Techniques*. Amsterdam: Morgan Kaufmann, 2011.

[31] M. Hall*, et al.* (2009) The WEKA data mining software: An update. *SIGKDD Explorations*.

[32] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics,* vol. 74, pp. 47-97, 2002.