

RESEARCH ARTICLE

# A decision support model for investment on P2P lending platform

Xiangxiang Zeng<sup>1</sup>, Li Liu<sup>1</sup>, Stephen Leung<sup>2</sup>, Jiangze Du<sup>3</sup>, Xun Wang<sup>4\*</sup>, Tao Li<sup>5\*</sup>

**1** School of Information Science and Technology, Xiamen University, Xiamen, China, **2** Faculty of Engineering, University of Hong Kong, Hong Kong, China, **3** School of Finance, Jiangxi University of Finance and Economics, Nanchang, China, **4** College of Computer and Communication of Engineering, China University of Petroleum, Qingdao, China, **5** School of Computer Science, Florida International University, Miami, FL, United States of America

\* wangsyun@upc.edu.cn (XW); taoli@cs.fiu.edu (TL)



**OPEN ACCESS**

**Citation:** Zeng X, Liu L, Leung S, Du J, Wang X, Li T (2017) A decision support model for investment on P2P lending platform. PLoS ONE 12(9): e0184242. <https://doi.org/10.1371/journal.pone.0184242>

**Editor:** Leyi Wei, Tianjin University, CHINA

**Received:** May 12, 2017

**Accepted:** August 18, 2017

**Published:** September 6, 2017

**Copyright:** © 2017 Zeng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data used in this study are third-party data (i.e. the authors did not collect the data and the authors do not claim ownership of the data). All data files are available from the website <https://www.kaggle.com/jschness/prosperloans>.

**Funding:** This work was supported by National Natural Science Foundation of China (61202011 and 61272152), and Ph.D. Programs Foundation of Ministry of Education of China (20120121120039).

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Peer-to-peer (P2P) lending, as a novel economic lending model, has triggered new challenges on making effective investment decisions. In a P2P lending platform, one lender can invest  $N$  loans and a loan may be accepted by  $M$  investors, thus forming a bipartite graph. Basing on the bipartite graph model, we built an iteration computation model to evaluate the unknown loans. To validate the proposed model, we perform extensive experiments on real-world data from the largest American P2P lending marketplace—Prosper. By comparing our experimental results with those obtained by Bayes and Logistic Regression, we show that our computation model can help borrowers select good loans and help lenders make good investment decisions. Experimental results also show that the Logistic classification model is a good complement to our iterative computation model, which motivates us to integrate the two classification models. The experimental results of the hybrid classification model demonstrate that the logistic classification model and our iteration computation model are complementary to each other. We conclude that the hybrid model (i.e., the integration of iterative computation model and Logistic classification model) is more efficient and stable than the individual model alone.

## Introduction

Peer-to-peer (P2P) lending is an emerging financial market. In recent years, more people are engaged in this financial platform. For example, the volume of business and the turnover of Prosper and Lending Club, the large-scale online P2P lending intermediary agents in the United States, are 50 million and nearly 100 million per month, respectively. The P2P lending is developing rapidly all over the world. Examples of P2P lending include PPDai in China and Zopa in Europe. In Prosper, individuals either request to borrow money, take a borrower role, or buy loans as a lender. The borrower sets the amount of money he or she needs and the maximum rate he or she would be willing to pay for this loan by posting a listing. Each lender will scan these loans to bid a partial amount and give the minimum rate they are willing to receive. The main difference between P2P lending and traditional bank industry is that in the former,

each lender can not only obtain the loan's financial information, but also evaluate the risk of bidding according to the borrower's social characteristic. On the other hand, Prosper, as an intermediary agent of P2P lending, collects many borrowers and lenders, help users borrow money quickly or gain benefits by investing. In this Internet platform, investors and borrowers form an  $M$ -to- $N$  relation model called bipartite graph, in which a lender can invest  $N$  loans, and a loan may be accepted by  $M$  investors.

P2P lending, as a burgeoning financial market, becomes a new field for academic research. In recent years, the social networking services on P2P lending have been explored extensively. Berger and Gleisner [1] found that these market participants act as financial intermediary and can significantly improve borrowers' credit conditions by reducing information asymmetries, predominantly for borrowers with less attractive risk characteristics. Freedman and Jin [2] examined what information problems exist on Prosper and whether social networks can help alleviate the information problems. They found that the estimated returns of groups loans are significantly lower than those of non-group loans partially due to lender learning and partially due to Prosper eliminating group leader rewards. Lin et al. [3] tested whether social networks lead to better lending outcomes, focusing on the distinction between the structural and relational network measures that are associated with a higher likelihood of a loan being funded, a lower risk of default, and lower interest rates. The social networks approaches were also applied in other fields, such as bioinformatics [4]. Collier and Hampshire [5] draw on the theory from the Principle-Agent perspective to empirically examine the signals that enhance community reputation. Sergio [6] measured the influence of social interactions in the risk evaluation of a money request; with a special focus on the impact of one-to-one and one-to-many relationships. His results showed that fostering social features increases the chances of getting a loan fully funded, when financial features are not sufficient to construct a differentiating successful credit request. Chen et al. [7] described an approach to measure the entrepreneurship orientation of online P2P lending platforms.

Some researchers take the perspective of borrowers when developing a model. Wu and Xu [8] proposed a decision support system based on intelligent agents in P2P lending for borrowers. The system provides borrowers with individual risk assessment, eligible lender search, lending combination and loan recommendation. The empirical results in [9] displayed that borrower' decisions, such as loan amount and interest rate, will determine whether he or she could successfully find loans or not. Herzenstein et al. [10] gave specific suggestions to borrowers to increase their chances of receiving funding in P2P lending communities. Pope and Sydnor [11] discovered that loan listings with blacks in the attached picture are 25% to 35% less likely to receive funding than those of whites with similar credit profiles.

To help investors make better investment decision, Luo et al. [12] proposed a data driven investment decision-making framework, which exploits the investor composition of each investment for enhancing decisions in P2P Lending. Katherine and Herrero-Lopez [13] examined the behavior of lender in a large peer-to-peer lending network and find that, while there exists high variance in risk-taking between individuals, many transactions represent sub-optimal decisions on the part of lenders. Lauri et al. [14] introduced a Borrower Decision Aid which helps formalize the decision making process of the sellers, or borrowers. Singh et al. [15] focused on risk and return of investments on Prosper. They found that within each credit grade, there exist subgroups which give positive return. For these subgroups, risk is aligned with return. In addition, the groups of loans with lower credit grades are more efficient in terms of risk and return alignment than those with higher credit grades. Klafft [16] demonstrated that following some simple investment rules may improve profitability of a portfolio and lead to acceptable returns for all credit rating categories with an exception of the high-risk ones. Garman et al. [17] introduced the concept of a research premium, which is the difference

between the interest rate borrowers would pay in a market with costless search and the minimum interest rate they must offer to induce search. Iyer et al. [18] evaluated whether lenders in such peer-to-peer markets are able to use borrower information to infer creditworthiness. They examined this capability using a methodology that takes advantage of the category of the borrowers and found that lenders are able to use available information to infer creditworthiness that is captured by a borrower's credit score.

P2P lending involves diverse elements, which renders research opportunities and challenges that are under-explored currently. The goal of this paper is to predict new investment abilities of investors and filter good requests in new loans on the basis of the bipartite graph model. This work focuses on synthesizing those old manifestations of investors in the past investments and providing a comprehensive evaluation. The performance of old investors' will help discover and analyze good new loans and reliable new investments according to the lending and investing model generated from P2P lending data.

The organization of this paper is as follows: Section 2 defines a  $M$ -to- $N$  relation model called bipartite graph, which is based on the relation of investors and loans in P2P lending. Section 3 quantitatively analyzes the comprehensive evaluation of new investors and unknown status loans by modeling an iteration computation model. Section 4 gives an integrated decision model to help investors pick trustworthy loans. Section 5 validates the effectiveness of the proposed investment decision model on the basis of real world data from the Prosper platform. By comparing current experimental results with those obtained by BayesNet [19], Logistic [20], and Average, we show that our computation model can help investors make better investment decisions. In addition, our empirical results demonstrate that the logistic classification model and the proposed iteration computation model complement each other. As a result, the hybrid model (i.e., the integration of our model and Logistic model) is more efficient and stable than each of the individual model. Finally, Section 6 presents conclusions of this work.

## Comprehensive evaluation of investor and load

### Investing network

P2P lending on Prosper is similar to that in the stock market, in which each investor can disperse his or her money to numerous loans and relatively one loan may be allocated by a multi-lender. This many-to-many relationship can be modeled as a *bipartite graph* which is widely used to model the relationship between two types of entities. We will use the *bipartite graph* to explain the relationship of investors and investees.

In the bipartite graph of P2P lending relationship, investors and borrowers are different types of entities, and the weight of edges is the amount of investors bidding for loan. In particular, on the Prosper online platform, if one borrower needs to borrow a sum of money, he or she must apply for a loan. After the application is approved by Prosper, a listing will be posted for all investors. Each lender who scans the loan can bid a partial amount until completed in full.

We use  $L = \{l_1, l_2, \dots, l_n\}$  to represent the set of  $N$  investors, and  $B = \{b_1, b_2, \dots, b_m\}$  stands for  $M$  borrowers.  $S = \{status_1, status_2, \dots, status_m\}$  represents the status of  $M$  loans and  $status_m = \{1, 0\}$ .  $e_{ij}$  is the amount that investor  $l_i$  has lent to borrower  $b_j$  and is equal to 0 if no correlation exists. Thus,  $E = \{e_{11}, e_{12}, \dots, e_{ij}, \dots, e_{nm}\}$  is the set of all biddings. Finally,  $L$ ,  $B$ , and  $E$  make up a bipartite investment network  $G = \{L, B, E\}$ . Fig 1 shows an example of this relation.

### Initial evaluation model

Basing on the bipartite graph of the investment network, we can propose a comprehensive evaluation model for investors and borrowers.

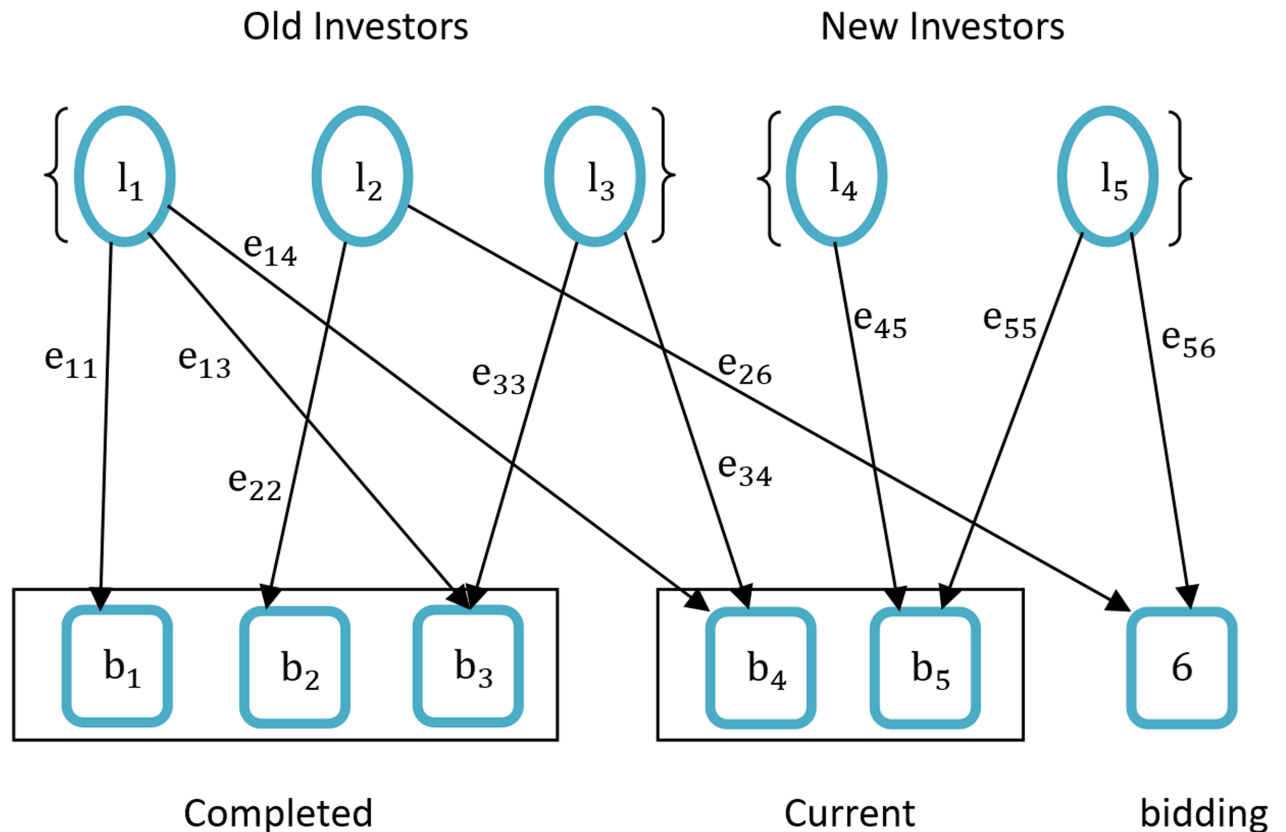


Fig 1. Simple bipartite graph of investment network.

<https://doi.org/10.1371/journal.pone.0184242.g001>

First, to judge whether a lender is a reliable investor, we need to determine his or her investment performance, which is equal to the investment amount and income that serve as moderator attributes. We use  $LS_i$  to stand for the confidence level of lender  $l_i$ . Note that  $status_j = \{0,1\}$  shows the status of loan  $b_j$ . Therefore,  $LS_i$  is the ratio of investor  $l_i$ 's total amount of paid investment loans to the total amount of investment loans in the investment network. So  $LS_i$  can be written as Eq (1):

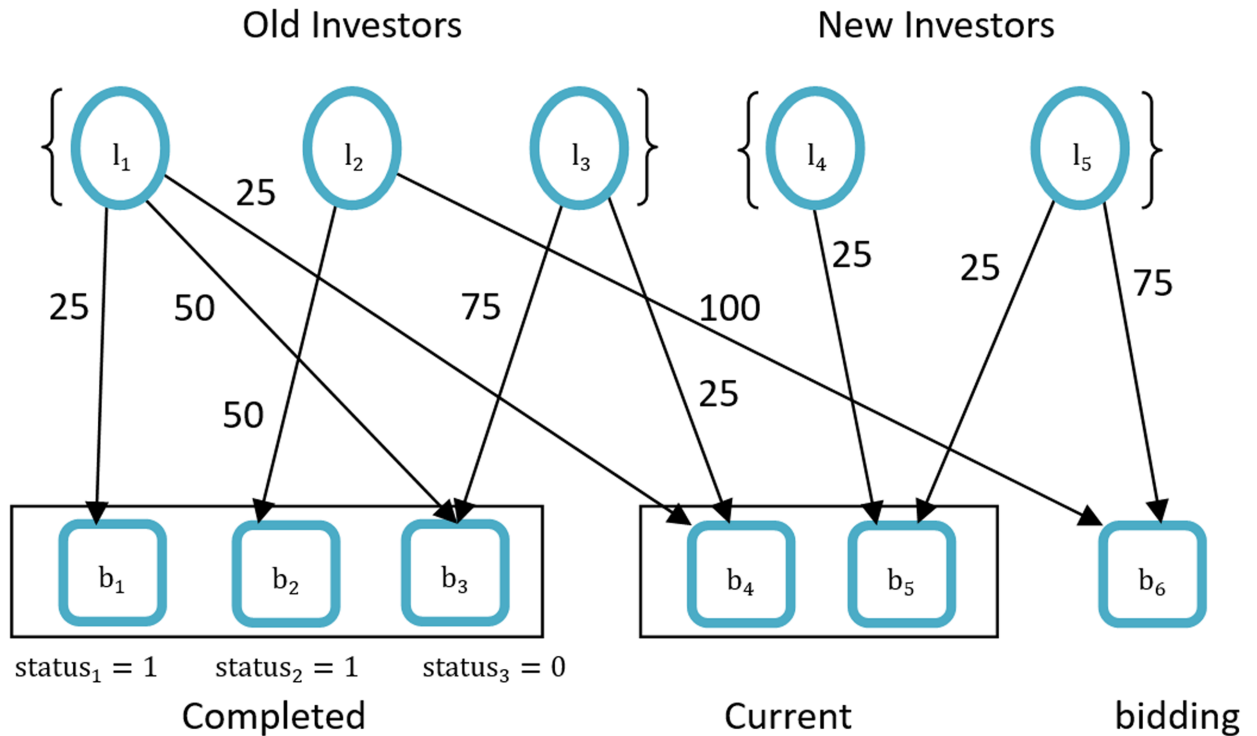
$$LS_i = \frac{\sum_{j=1}^m e_{ij} * status_j}{\sum_{j=1}^m e_{ij}}, i = 1, 2, \dots, n \tag{1}$$

where  $m$  is the number of completed loans and  $n$  is the number of old investors.

When investors who have finished some biddings are evaluated, we can rely on their confidence level to analyze the current and bidding loans to which they have allocated money. We use  $BS_j$  to stand for the predicted paid probability of the current or bidding loan  $b_j$ , which can be computed by Eq (2):

$$BS_j = \frac{\sum_{i=1}^n e_{ij} * LS_i}{\sum_{i=1}^n e_{ij}}, j = 1, 2, \dots, m \tag{2}$$

where  $n$  represents the quantity of investors and  $m$  is the quantity of current and bidding loans.



Old Investors :  $LS_1 = \frac{25 \cdot 1 + 50 \cdot 0}{25 + 50} = 0.33, LS_2 = \frac{50 \cdot 1}{50} = 1, LS_3 = \frac{75 \cdot 0}{75} = 0$

Current and New Loans :  $BS_4 = \frac{25 \cdot 0.33 + 25 \cdot 0}{25 + 25} = 0.17, BS_6 = \frac{100 \cdot 1}{100} = 1$

New Investors :  $LS_5 = \frac{75 \cdot 1}{75} = 1$

Fig 2. Initial evaluation.

<https://doi.org/10.1371/journal.pone.0184242.g002>

Basing on the paid probability of the current or bidding loans, we will evaluate the confidence level of novices, as written in Eq (3):

$$new LS_i = \frac{\sum_{j=1}^n e_{ij} * BS_j}{\sum_{j=1}^n e_{ij}}, i = 1, 2, \dots, m \tag{3}$$

where  $n$  is the number of current and bidding loans, and  $m$  is the number of new investors. A simple case of initial evaluation is given in Fig 2.

A problem exists with this initial evaluation. In the above case, we can observe that the confidence level  $LS_4$  of investor  $l_4$  and the paid probability  $BS_5$  of loan  $b_5$  are not evaluated. However, by carefully analyzing the result and the bipartite graph of the investment network, we learnt that  $BS_5$  can be evaluated by the relationship with investor  $l_5$ . Similarly, now that  $BS_5$  is known,  $LS_4$  can also be calculated according to the relationship with loan  $b_5$ . Therefore, we need to conduct a second round of calculation base on initial computation results.

## Second round evaluation

Some current and bidding loans evaluated from the first round of calculation are evaluated in the second round of evaluation. Therefore, to evaluate the old investor's  $LS'_i$  again, we should employ the paid, current, and bidding loans:

$$LS'_i = \frac{\sum_{j=1}^m e_{ij} * status_j + \sum_{j=m+1}^M e_{ij} * BS_j}{\sum_{j=1}^M e_{ij}}, i = 1, 2, \dots, n \tag{4}$$

where  $m$  is the number of completed loans,  $M$  is the number of all loans and  $n$  is the number of all investors. Therefore, Eq (3) is not needed in the calculation, and the current and bidding loans' paid probability  $BS'_j$  in the second round can be written as:

$$BS'_j = \frac{\sum_{i=1}^n e_{ij} * LS'_i}{\sum_{i=1}^n e_{ij}}, j = 1, 2, \dots, m \tag{5}$$

where  $n$  represents the quantity of investors and  $m$  is the quantity of current and bidding loans. Therefore, by using Eq (5),  $BS'_5 = 25 * 1/25 = 1$ . The integrated results are shown in Fig 3.

## Multi-round and convergence

Unfortunately, investor  $l_4$  is still not correlated and evaluated. Nevertheless, this problem is not serious, because we can easily predict the value of  $l_4$  by executing Eq (4) again. We obtain  $LS'_4 = 25 * 1/25 = 1$ . Fig 4 shows the results of the investors' confidence level in the third round. Eqs (4) and (5) are recursive. Therefore they can be computed with initial states and by iterating the computation until convergence. Fig 5 demonstrates a consistent steady state solution for an investment network.

## Convergence properties

In Fig 5, all investors and loans converge to a steady state. In this work, we use real data from Prosper that includes more than 300,000 bids, more than 4,000 loans, and more than 12,000 investors. The evaluation process takes 25 iterations to converge. Fig 6 shows the rate of convergence with different numbers of loans. The graph suggests that our computation model scale well even for large collections as the scaling factor is roughly linear in  $\log n$ . This computation process is similar to the PageRank [21] computation with the convergence property as described in [22]. Our model computation terminates in logarithmic time, which is equivalent to saying that the bipartite graph has a good expansion factor.

## Decision model

In the P2P lending market on Prosper, there are many loans that are posted at same time; and we need to assess those loans whether they will be paid. In the stock market, smart investors usually make good selections (i.e., wise decisions). Similarly, in the P2P lending market, investors have different knowledge backgrounds and may make good or bad decisions. Therefore, good investors are our reference objects before making a decision.

## Model description

Fig 7 describes the detailed process of computation and decision. This process can be divided into four parts. First, in the process from line 1 to line 3, we need to input the training data  $DataSetT$  and the predicted data set  $DataSetP$ , which contain the most important information

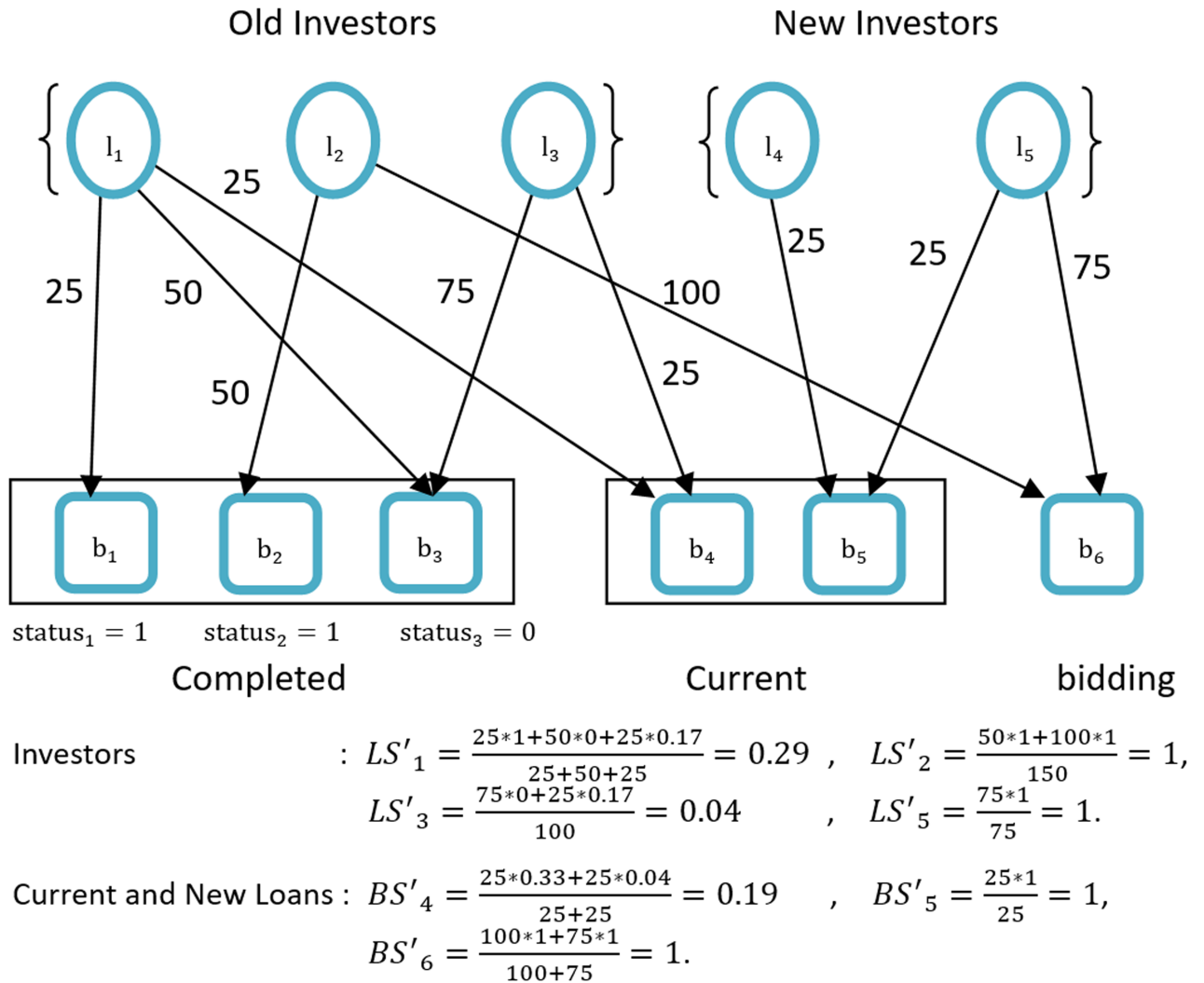


Fig 3. Second evaluation.

<https://doi.org/10.1371/journal.pone.0184242.g003>

(such as bids and loans). The bid table records each investor’s bidding to a loan, one line for each loan. The loan table stores recorded unique loans We use the two types of table to build the *bipartite graph* which depicts the correlation of investors and loans. We also employ these tables to obtain the computation data of  $L, B, E, S$ , which correspond to the set of investors, borrowers, bidding, and loan status, respectively. Second, initial evaluation begins at line 4. We use Eq (1) to compute the confidence level of all old investors in line 1 to line 3 of the initial evaluation, Eq (2) to predict the paid probability of all new loans in line 4 to line 6, and Eq (3) to estimate the confidence level of new investors in line 7 to line 9. Third, in line 5, the goal of multi-round iteration is to evaluate the underlying confidence level of investors and the paid probability of loans until convergence. We will iterate this process until Total Difference from Previous Iteration (TDPI) converge to  $\lambda$  which is set at a reasonable tolerance value. In line 3 to line 5 of Multi-Round and Convergence, we utilize Eq (4) to calculate the confidence level of all investors and employ Eq (5) to estimate the paid probability of all new loans in line 6 to line 8. Finally, we sort the paid probability of all current or bidding loans and generate the result candidate set RCS that are the top  $\alpha$  ranking based on the paid probability.

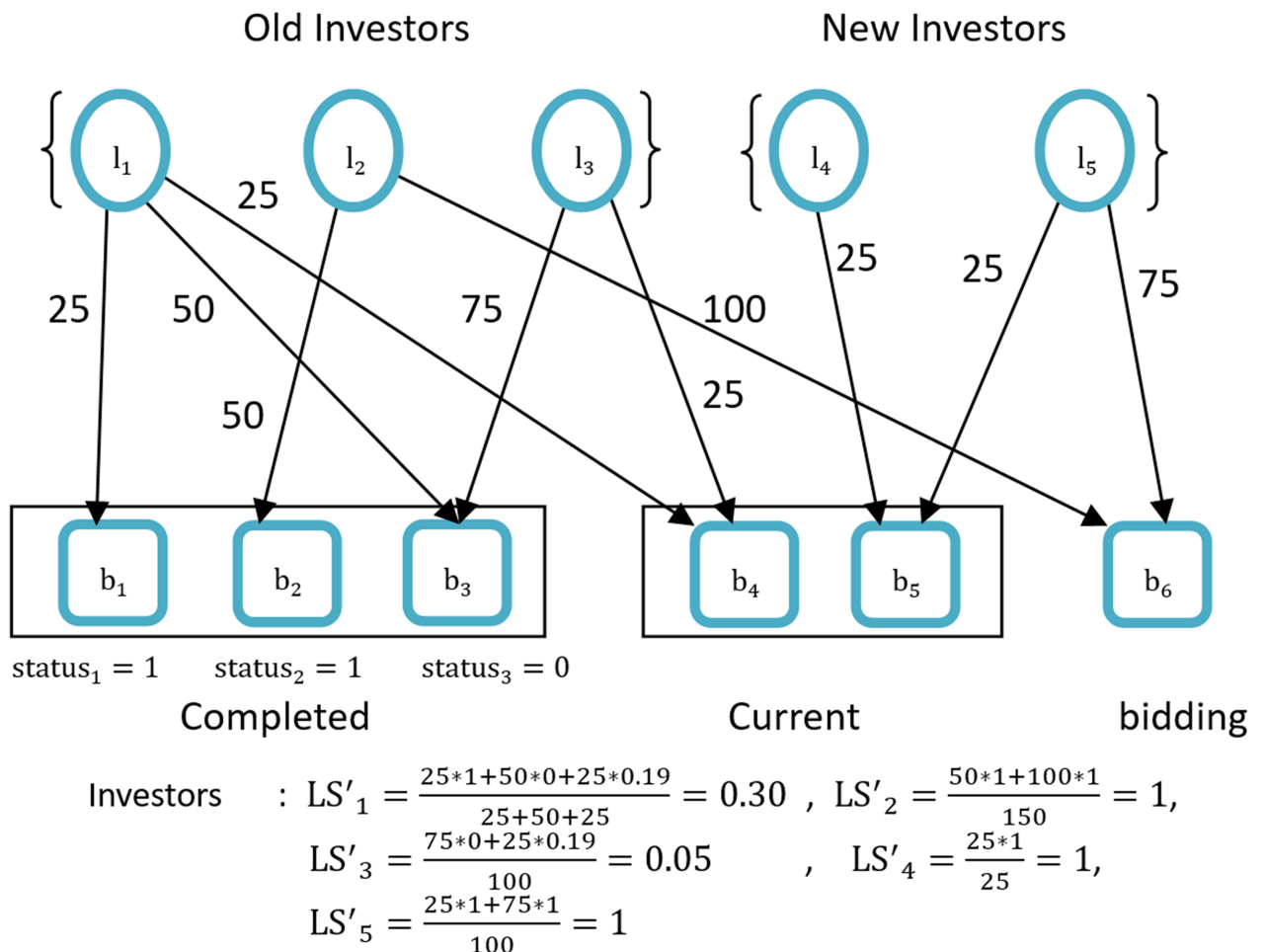
### Experimental results

In this section, we validate our model by employing real-world data on P2P lend platform. We first describe the data that are used in the experiments. In the following sections, we illustrate the effectiveness of the results predicted by the proposed model. We find that our decision model can effectively locate the paid loans.

### Experimental data description

In our experiment, we download the data from Prosper, and select a training data set from January 2009 to June 2010 and a predicted data set from January 2010 to June 2010. We have selected Prosper because it is deemed relatively stable in this period. We seek two types of the most important tables, namely bids and loans to train and test our model.

The bids table contains MemberKey, ListingKey, Amount, and CreationDate. These fields represent unique investor, unique loan, amount of one investor bidding for a loan, and the bid creation date, respectively. A bid is created when a lender wishes to lend money to a borrower. The loans table includes ListingKey, Status, and CreationDate. These fields stand for unique loan, loan status (i.e., current, paid or unpaid), and creation time of the loan, respectively. These information and correlation are useful in building the bipartite network of the relationship of investors and loans.



**Fig 4. Investor confidence level in the third round.**

<https://doi.org/10.1371/journal.pone.0184242.g004>



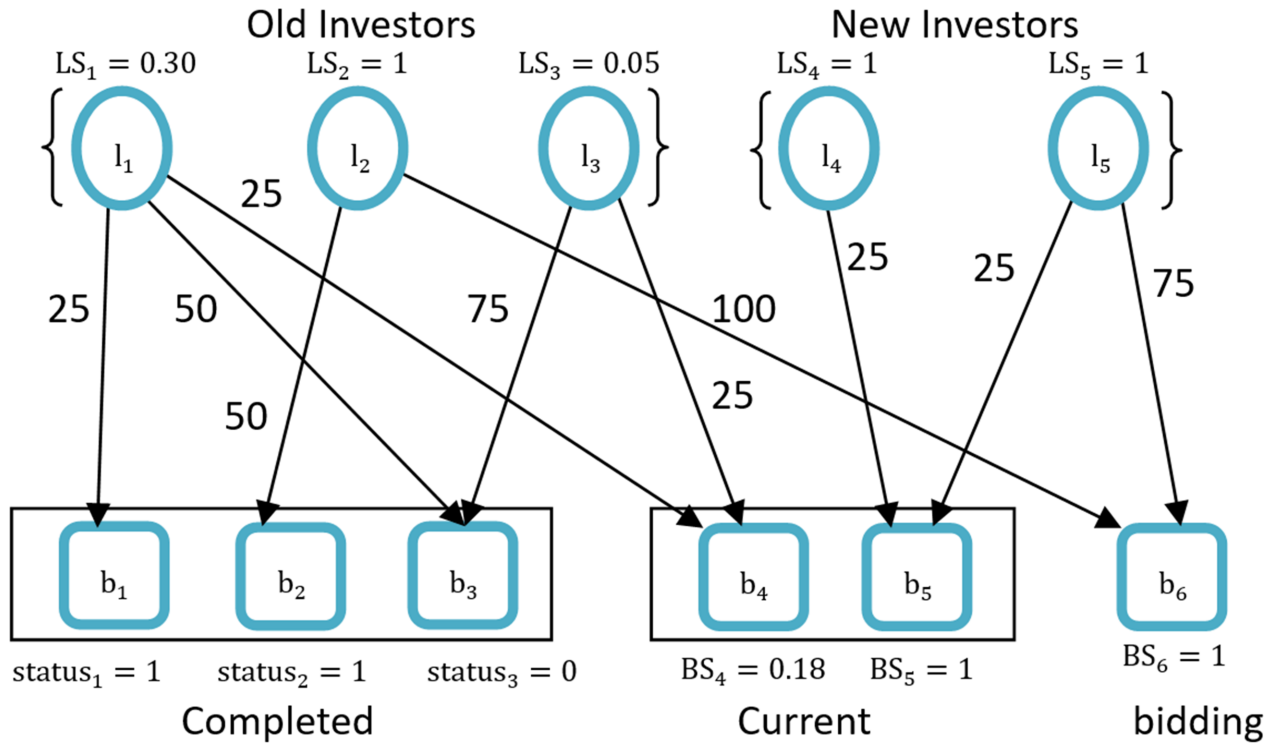


Fig 5. Convergence state of a simple case.

<https://doi.org/10.1371/journal.pone.0184242.g005>

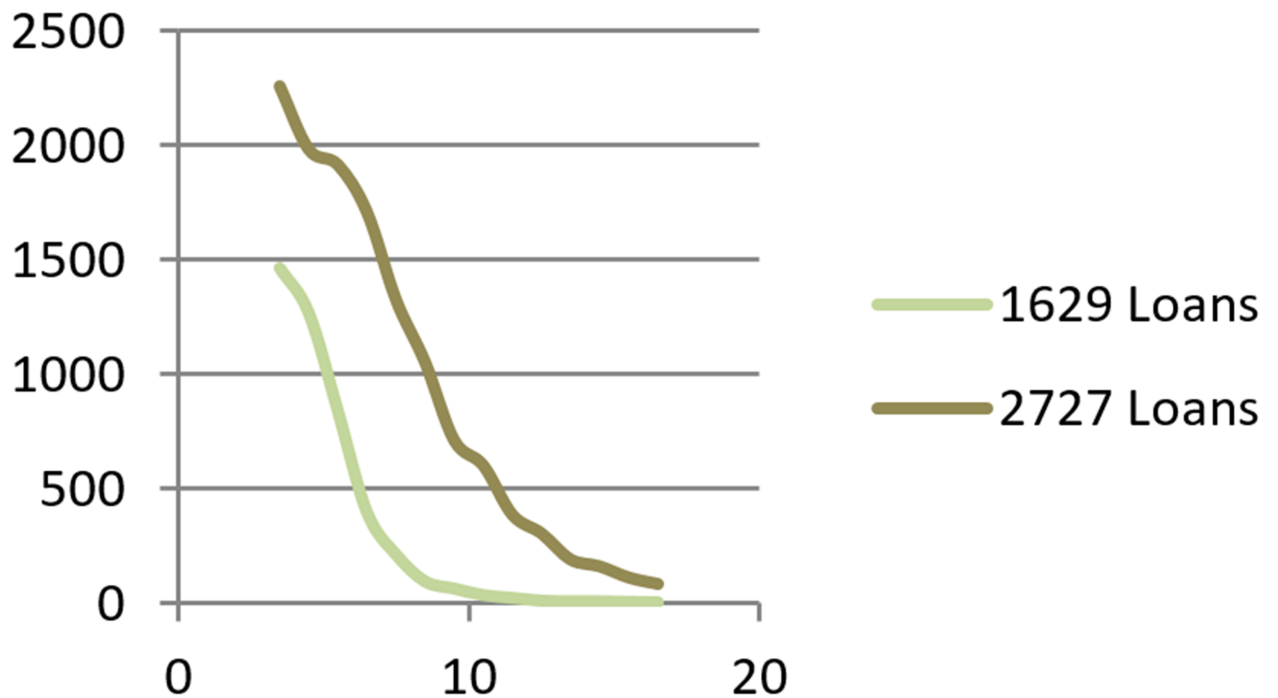


Fig 6. Rate of convergence with different numbers of loans.

<https://doi.org/10.1371/journal.pone.0184242.g006>

1, Input training data set DataSetT and predicted data set DataSetP

2, Select Bids, Loans table from DataSetT and DataSetP

3,  $(L, B, S, E) = \text{BuildNet}(\text{Bid}, \text{Loan})$

4, Initial Evaluation

(1) For  $i = 1$  to the Number of All Old Investors

$$(2) \quad LS_i = \frac{\sum_{j=1}^m e_{ij} * status_j}{\sum_{j=1}^m e_{ij}}$$

(3) End For

(4) For  $j = 1$  to the Number of All New Loans

$$(5) \quad BS_j = \frac{\sum_{i=1}^n e_{ij} * LS_i}{\sum_{i=1}^n e_{ij}}$$

(6) End For

(7) For  $i = 1$  to the Number of New Investors

$$(8) \quad new \ LS_i = \frac{\sum_{j=1}^n e_{ij} * BS_j}{\sum_{j=1}^n e_{ij}}$$

(9) End For

5, Multi-Round and Convergence

(1) TDPI = the Number of All New Loans

(2) For  $TDPI > \lambda$

(3) For  $i = 1$  to the Number of All Investors

$$(4) \quad LS'_i = \frac{\sum_{j=1}^m e_{ij} * status_j + \sum_{j=m+1}^M e_{ij} * BS_j}{\sum_{j=1}^M e_{ij}}$$

(5) End For

(6) For  $j = 1$  to the Number of All New Loans

$$(7) \quad BS'_j = \frac{\sum_{i=1}^n e_{ij} * LS'_i}{\sum_{i=1}^n e_{ij}}$$

(8) End For

(9)  $TDPI = \text{TotalDifferencefromPreviousIteration}(BS)$

(10) End For

6, Gain Investment Candidate Set

(1)  $CS = \text{TopRanking}(BS)$

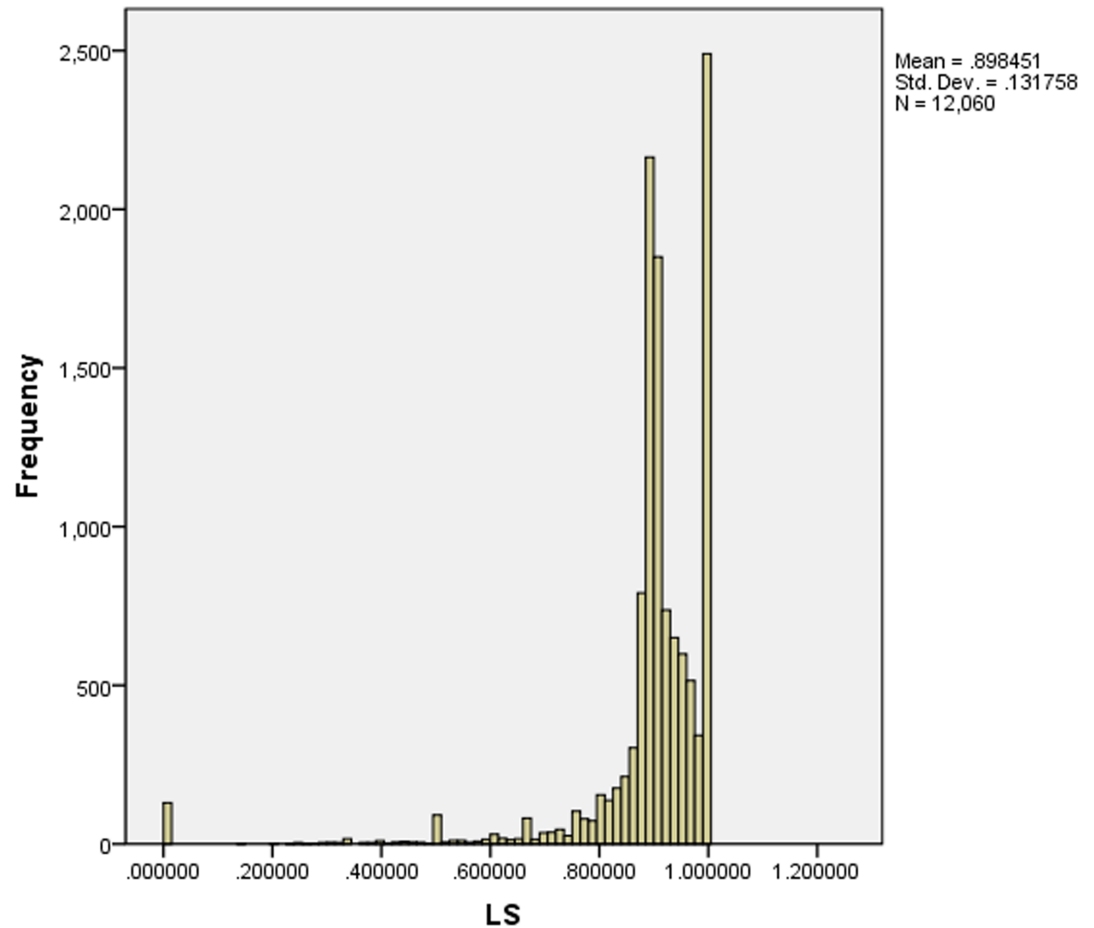
(2)  $RCS = \text{ResCandiSet}(CS, \alpha)$

7, Output RCS

**Fig 7. Investment decision process.**

<https://doi.org/10.1371/journal.pone.0184242.g007>

In our selection data set, the bids table recorded 300,384 recorders and loans table saves 4,335 loans.



**Fig 8. Confidence level of investors.**

<https://doi.org/10.1371/journal.pone.0184242.g008>

### Result analysis

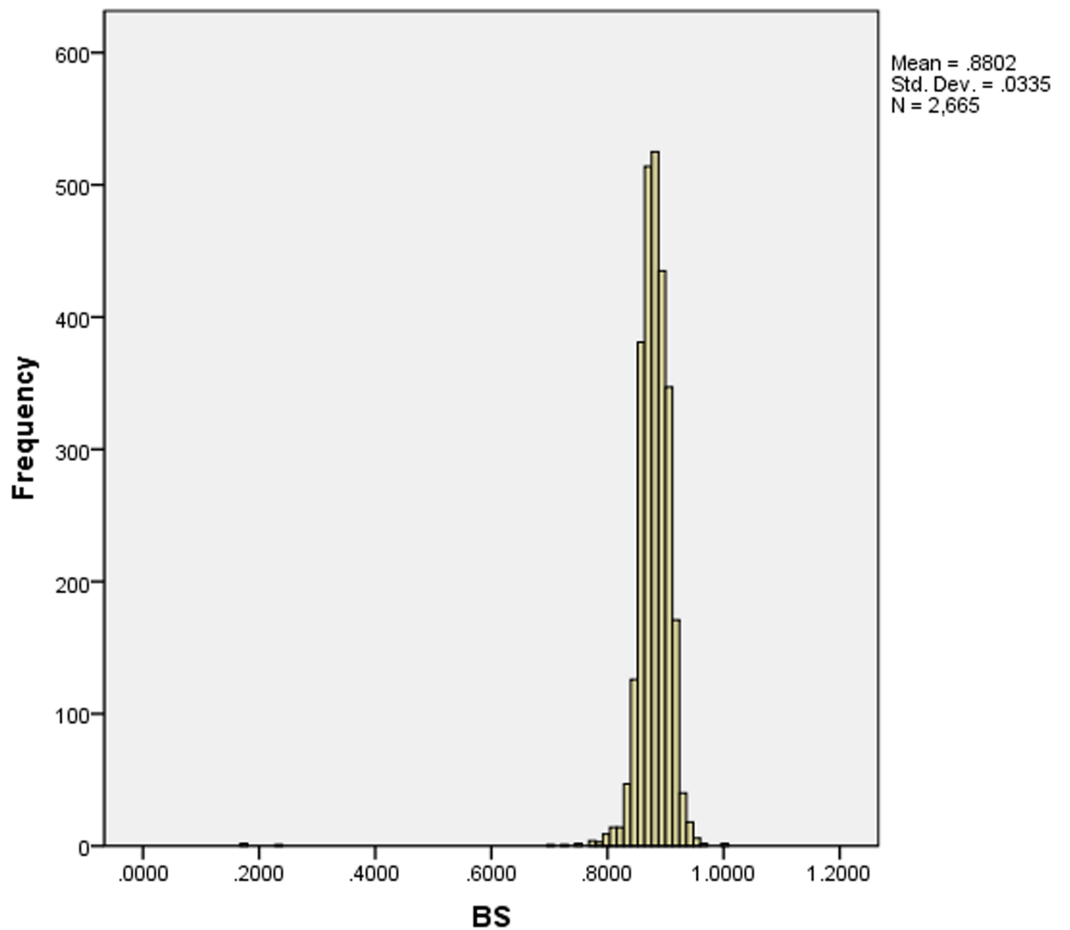
In the following section, we illustrate how our decision model is validated to predict whether a loan will be paid or unpaid by analyzing computation results. Fig 8 shows the confidence level of the investors’ statistical histogram, whose results converge after 24 iterations. This chart suggests that the confidence level of the majority of investors is less than 0.9 and that a considerable amount of investors have good performance. We employ the confidence level of investors *LS* to predict the paid probability. Fig 9 shows the statistical histogram of the paid probability of loans.

Fig 9 shows that the paid probability of loans presents a Gaussian distribution. Additionally, we utilize Eq (6) to calculate the accuracy rate *AR* at given interval  $\theta$ . Results are given in Table 1.

$$AR = \frac{N_1}{N_1 + N_0} \tag{6}$$

Here,  $N_1$  is the number of paid loan, and  $N_0$  is the number of unpaid loan.

Table 1 shows that the accuracy rate increases when the degree of paid probability interval increases. Fig 10 illustrates that the unpaid number significantly decreases when the degree of paid probability increases. Therefore, our decision model is effective to help investors select



**Fig 9. Paid probability of loans.**

<https://doi.org/10.1371/journal.pone.0184242.g009>

loans. The results also indicates that some potential benefits may be lost if we assign  $\theta$  a very large value.

Fig 11 shows that 50% of paid probability is higher than the unpaid. This result is a good sign that validates our decision model.

### Result comparisons

In P2P lending, our decision model identifies the ranking of loans, from the best loans to the worse loans, according to the paid probability of loan. Investors can then set a customized definition variable  $\alpha$  to select the top set as the candidate investment set. Fig 12, compares the paid rate. This figure uses the average paid rate as the baseline. The chart shows that investors make investment decisions in a random way. Approximately 88.3% of loans can be paid, which

**Table 1. Accuracy rate of paid probability interval.**

$\theta$	<0.8	0.8to0.82	0.82to0.84	0.84to0.86	0.86to0.88	0.88to0.9	0.9to0.92	0.92to0.94	>0.94
Pain	5	13	63	322	484	456	328	51	14
Total	10	16	75	389	557	508	344	52	15
AR	0.5	0.81	0.84	0.83	0.87	0.90	0.95	0.98	0.93

<https://doi.org/10.1371/journal.pone.0184242.t001>

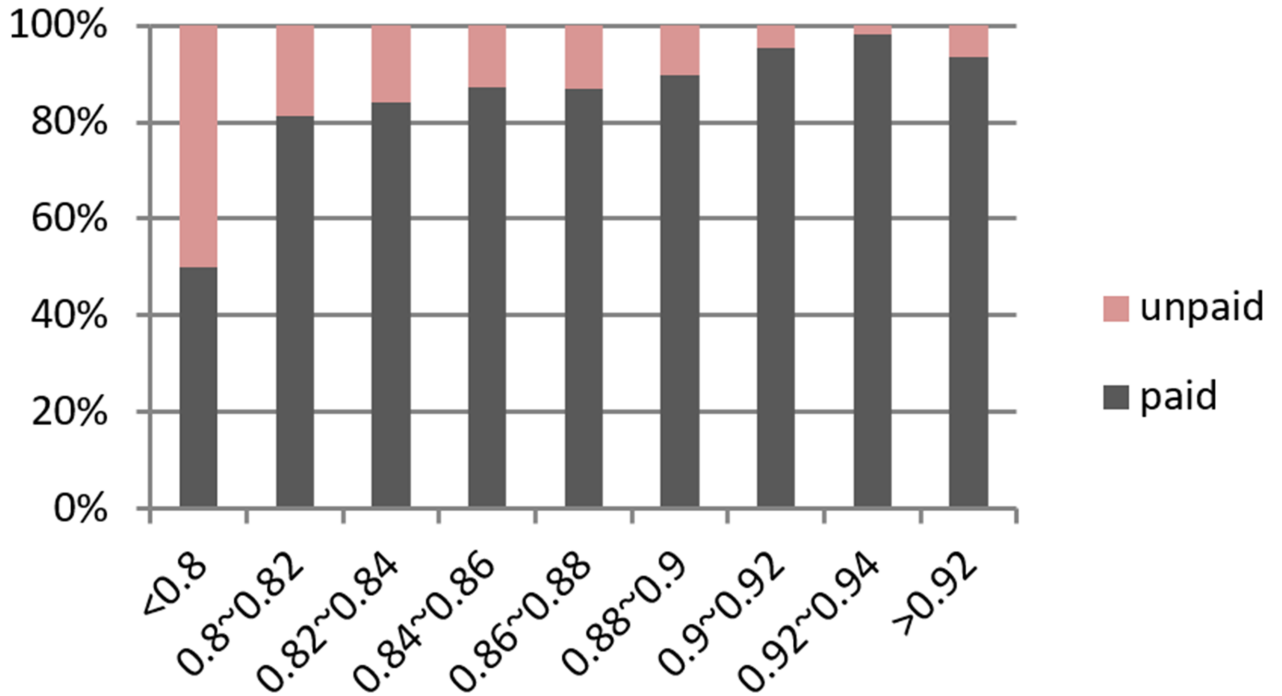


Fig 10. Contrast ratio of paid and unpaid probability of investors.

<https://doi.org/10.1371/journal.pone.0184242.g010>

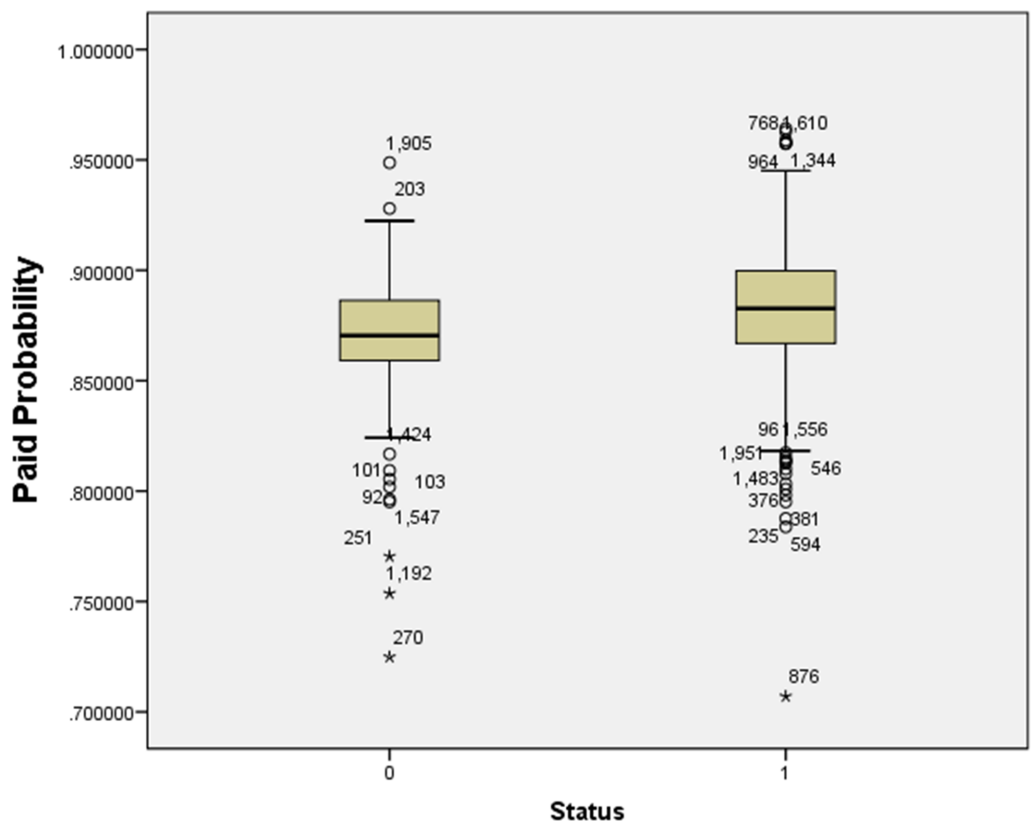
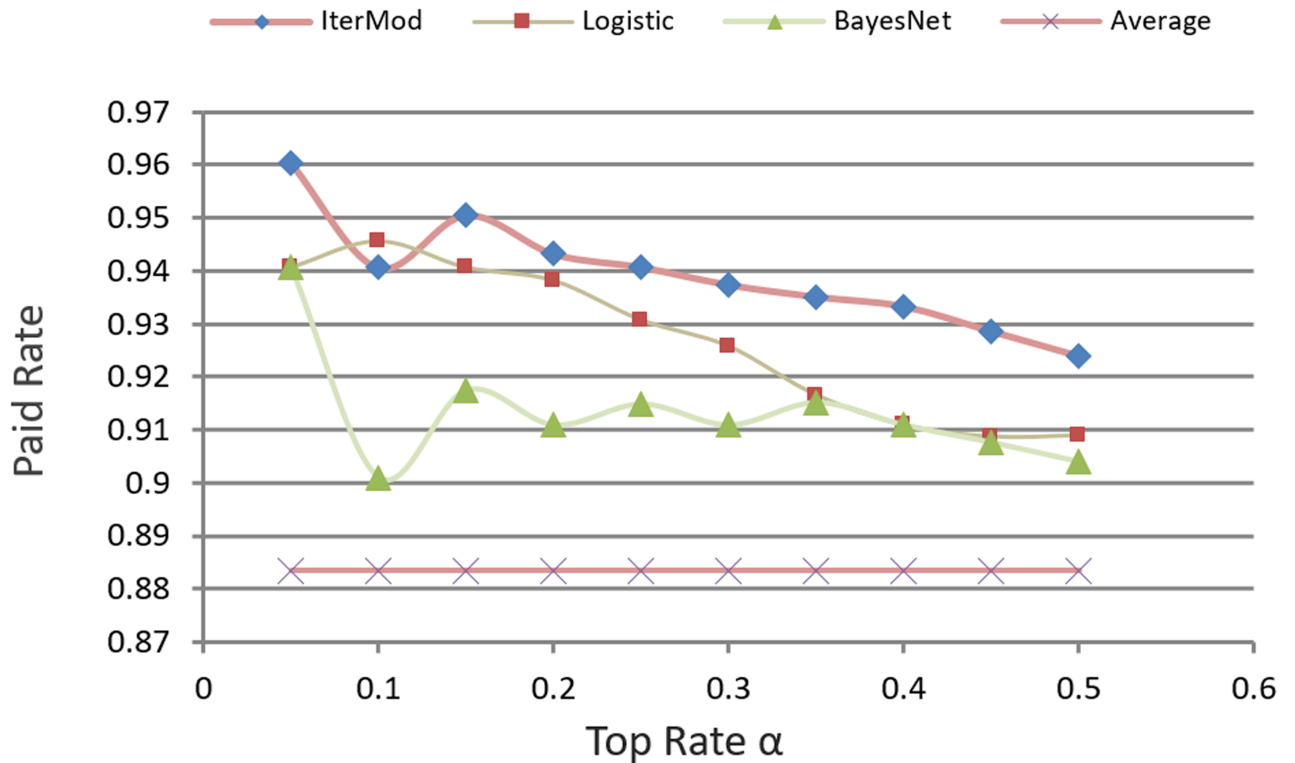


Fig 11. Box plot for paid and unpaid probability investors.

<https://doi.org/10.1371/journal.pone.0184242.g011>



**Fig 12. Comparison of paid rate.**

<https://doi.org/10.1371/journal.pone.0184242.g012>

shows that Prosper is doing a good job in monitoring the transaction process. We compare our candidate selection with logistic regression and BayesNet, which are the most commonly used methods on personal loans risk analysis. For Logistic Regression and BayesNet, loans are ranked in the decreasing order of the expected paid rate. We can explicitly see that the evaluation result of our decision model is more effective than that of Logistic Regression, BayesNet, and Average. Although a slight decline relative to Logistic Regression exists when  $\alpha = 0.1$ , the candidate set chosen by our model still has higher paid rate than that of others.

### IterMod and logistic hybrid

Fig 12 shows that BayesNet and IterMod are singular when  $\alpha$  equals 0.1. On the other hand, Logistic is more active than the other two. We hypothesize that Logistic will achieve better performance if mixed with the computation model of IterMod. To integrate the two classification models, we introduce Eq (7). Thus, the paid probability  $BS_i$  of loan  $b_i$  consists of its confidence level from IterMod and Logistic.

$$BS_i = \theta IBS_i + (1 - \theta)LBS_i \tag{7}$$

Here,  $\theta$  is the confidence weight that represents the confidence level for the result of IterMod.

Fig 13, compares the paid rate of the hybrid model of IterMod and Logistic with other models. Results show that the hybrid model of IterMod and Logistic is efficient and has a stable performance for making investment decisions. In this experiment, we set  $\theta$  as 0.6. This is because experimental results show that iterative computation model achieves better performance than Logistic. Based on this comparison result, we conclude that Logistic and our computation model are complementary to each other.

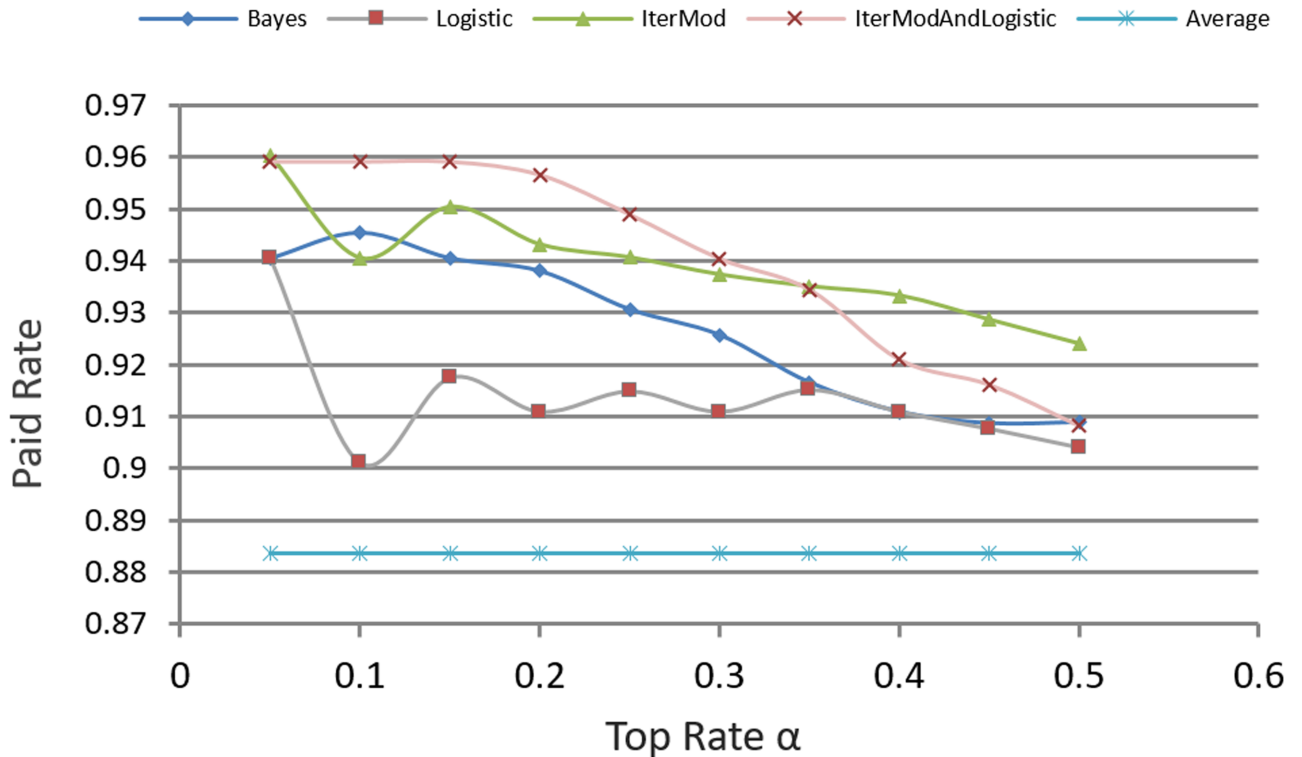


Fig 13. Comparison of paid rate of models.

<https://doi.org/10.1371/journal.pone.0184242.g013>

Fig 14 shows the comparison result when we set  $\theta = 0.5$ . The figure also verifies our hypothesis that the hybrid classification model can obtain better performance compared with other models. The most remarkable point is  $\alpha = 0.1$ , where IterMod's, BayesNet's, and Logistic's performance are singular, Logistic acts as an important complement in the hybrid model to improve the investment decision. Therefore, our computation model and Logistic complement each other, and the hybrid model by integrating both approaches is more efficient than each of the individual model alone.

### Concluding remarks

In this paper, we propose an investment decision model on P2P lending. We exploit the investment performance of old investors on the basis of their confidence level, predict new investors' confidence level, and estimate current or bidding loans' paid probability. Specifically, we employ historical records to compute the confidence level of old investors. Using the confidence level of old investors, we estimate the paid probability of new loans, which in return is utilized to predict the confidence level of new investors. Subsequently, we utilize the paid probability and the status of all historical and new loans to calculate the confidence level of all investors. This confidence level is then used to estimate the paid probability of all new loans. Finally, we iterate the process until the confidence level of all investors and all new loans converges. Experimental results on real-world P2P lending data demonstrate that our decision model can effectively filter out reliable borrowers and significantly improve the investment performance of investors. Experiments also show that the Logistic classification model and our iteration computation model complement each other, which inspired us to integrate the two classification models. Experiments show that the hybrid model (i.e., the integration of iterative

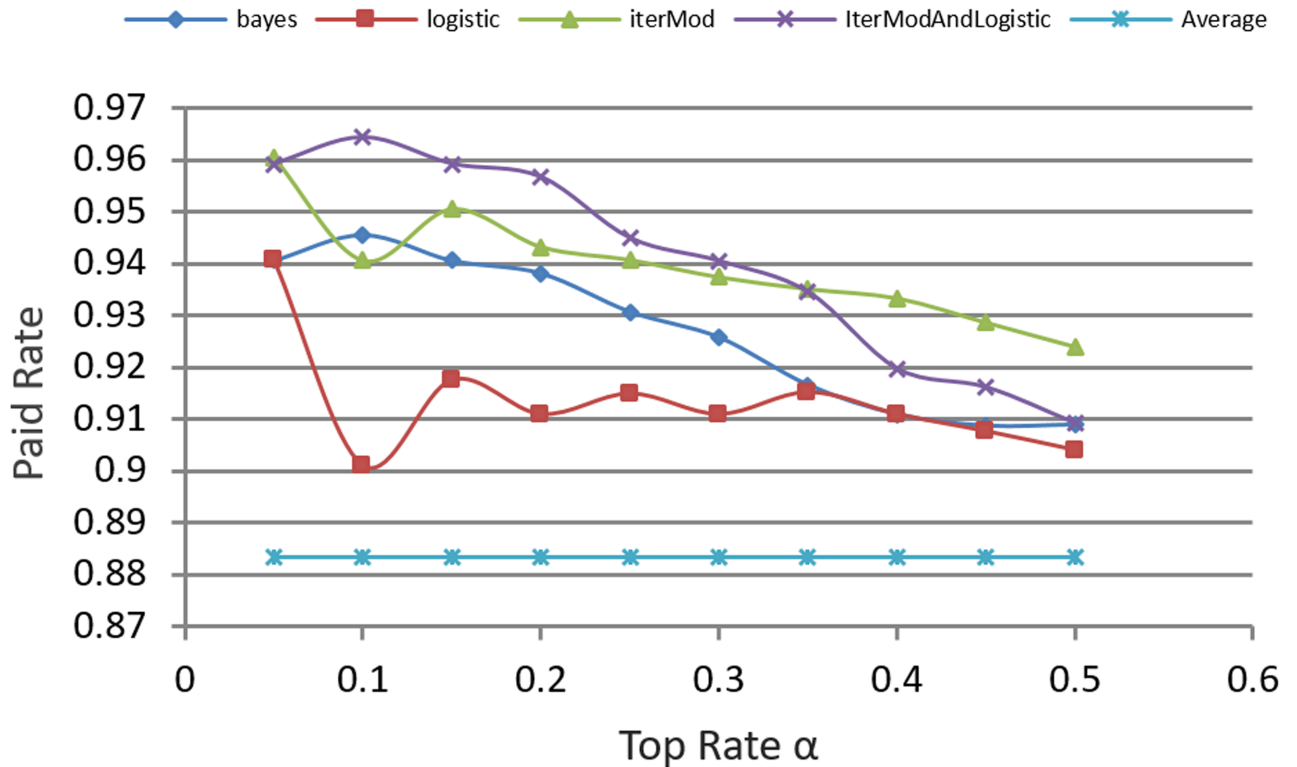


Fig 14. Comparison of paid rate of  $\theta = 0.5$ .

<https://doi.org/10.1371/journal.pone.0184242.g014>

computation model and Logistic classification model) is more efficient and stable than the individual model alone. In the future, an interesting work is to investigate P2P lending by considering multi-objective evolutionary algorithms (MOEAs) [23], since the effectiveness of MOEAs have been verified on a variety of real-world problems [24–26]. And machine learning classification model [27–29] such as the state-of-art  $v$ -support vector machine methods [30–32], dimension reduction techniques [33–35], neural-like computing models [36–38] and spiking neural networks [39, 40] could be tested. Community detection methods, especially overlapping communities detection [41] also can be considered for the further improvement.

### Author Contributions

**Conceptualization:** Xiangxiang Zeng.

**Data curation:** Li Liu, Jiangze Du.

**Formal analysis:** Li Liu.

**Funding acquisition:** Xiangxiang Zeng.

**Investigation:** Xiangxiang Zeng, Li Liu, Jiangze Du.

**Project administration:** Jiangze Du.

**Supervision:** Xiangxiang Zeng, Xun Wang, Tao Li.

**Validation:** Stephen Leung.

**Visualization:** Xun Wang.



**Writing – original draft:** Li Liu.

**Writing – review & editing:** Xiangxiang Zeng, Stephen Leung, Xun Wang, Tao Li.

## References

1. Berger SC, Gleisner F. Emergence of financial intermediaries in electronic markets: the case of online P2P lending. *Business Research*. 2009; 2(1):39–65.
2. Freedman S, Jin GZ. Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from [Prosper.com](http://Prosper.com). NET Institute Working Paper 2008;( 8–43).
3. Lin M, Prabhala NR, Viswanathan S. Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science*. 2013; 59(1):17–35.
4. Liu B, Chen J, Wang X. Application of Learning to Rank to protein remote homology detection. *Bioinformatics*. 2015; 31(21):3492–8. <https://doi.org/10.1093/bioinformatics/btv413> PMID: 26163693
5. Collier BC, Hampshire R. Sending mixed signals: multilevel reputation effects in peer-to-peer lending markets. Proceedings of the 2010 ACM conference on Computer supported cooperative work; Savannah, Georgia, USA. 1718955: ACM; 2010. p. 197–206.
6. Herrero-Lopez S, editor Social interactions in P2P lending. Proceeding SNA-KDD '09 Proceedings of the 3rd Workshop on Social Network Mining and Analysis; 2009 June 28–28, 2009; Paris, France
7. Chen X, Yang L, Wang P, Yue W. An Effective Interval-Valued Intuitionistic Fuzzy Entropy to Evaluate Entrepreneurship Orientation of Online P2P Lending Platforms. *Advances in Mathematical Physics*. 2013; 2013:8. <https://doi.org/10.1155/2013/467215>
8. Wu J, Xu Y. A Decision Support System for Borrower's Loan in P2P Lending. *Journal of Computers*. 2011; 6(6):1183–90.
9. Li S, Qiu J, Lin Z, Qiu J. Do borrowers make homogeneous decisions in online P2P lending market? An empirical study of PPDai in China. 2011:1–6.
10. Herzenstein M, Andrews RL. The Democratization of Personal Consumer Loans? Determinants of Success in Online Peer-To-Peer Loan Auctions. *Bulletin of the University of Delaware*. 2008; 15(3):274–7.
11. Pope DG, Sydnor JR. What's in a Picture? Evidence of Discrimination from [Prosper.com](http://Prosper.com). *Journal of Human Resources*. 2008; 46(1):53–92.
12. Luo C, Xiong H, Zhou W, Guo Y, Deng G. Enhancing investment decisions in P2P lending: an investor composition perspective. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining; San Diego, California, USA. 2020458: ACM; 2011. p. 292–300.
13. Krumme K, Herrero-Lopez S. Do Lenders Make Optimal Decisions in a Peer-to-Peer Network? Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology—Volume 01. 1632420: IEEE Computer Society; 2009. p. 124–7.
14. Puro L, Teich JE, Wallenius H, Wallenius J. Borrower Decision Aid for people-to-people lending. *Decision Support Systems*. 2010; 49(1):52–60.
15. Singh H, Gopal RD, Li X, editors. Risk and return of investments in online peer-to-peer lending. workshop on information technologies and systems(WITS 2009); 2009.
16. Klafft M, editor Online Peer-to-Peer Lending: A Lenders' Perspective. International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government, Eee 2008, July 14–17, 2008, Las Vegas, Nevada, Usa; 2008.
17. Samuel G, Robert H, Ramayya K. A Search Theoretic Model of Person-to-Person Lending. 2008.
18. Iyer R, Khwaja AI, Luttmer EFP, Shue K. Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? *Social Science Electronic Publishing*. 2009;15242(rwp09-031).
19. Wang X, Zhang D, Zeng X, Wu X. A Bayesian Investment Model for Online P2P Lending. In: Su J, Zhao B, Sun Z, Wang X, Wang F, Xu K, editors. *Frontiers in Internet Technologies: Second CCF Internet Conference of China, ICoC 2013, Zhangjiajie, China, July 10, 2013, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 21–30.
20. Dong G, Lai KK, Yen J. Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*. 2010; 1(1):2463–8.
21. Page L. The PageRank citation ranking: Bringing order to the web. Stanford Digital Libraries Working Paper. 1998; 9(1):1–14.
22. Motwani R, Raghavan P. Randomized Algorithms. *Acm Sigact News*. 1995; 26(3):48–50.

23. Zhang X, Tian Y, Cheng R, Jin Y. A Decision Variable Clustering Based Evolutionary Algorithm for Large-scale Many-objective Optimization. *IEEE Transactions on Evolutionary Computation*. 2016. <https://doi.org/10.1109/TEVC.2015.2454857>
24. Zhang X, Duan F, Zhang L, Cheng F, Jin Y, Tang K. Pattern Recommendation in Task-oriented Applications: A Multi-Objective Perspective. *IEEE Computational Intelligence Magazine*. 2017.
25. Zhang L, Pan H, Su Y, Zhang X, Niu Y. A Mixed Representation based Multi-Objective Evolutionary Algorithm for Overlapping Community Detection. *IEEE Transactions on Cybernetics*. 2017.
26. Xue Y, Jiang J, Zhao B, Ma T. A self-adaptive artificial bee colony algorithm based on global best for global optimization. *Soft Computing*. 2017. <https://doi.org/10.1007/s00500-017-2547-1>
27. Chen L, Chen W, Qiu C, Wu Y, Krishnan S, Zou Q. LibD3C: Ensemble classifiers with a clustering and dynamic selection strategy. *Neurocomputing*. 2014; 123:424–35.
28. Zou Q, Liu W, Merler M, Ji R. Advanced learning for large-scale heterogeneous computing. *Neurocomputing*. 2016; 217:1–2.
29. Liu B, Zhang D, Xu R, Xu J, Wang X, Chen Q, et al. Combining evolutionary information extracted from frequency profiles with sequence-based kernels for protein remote homology detection. *Bioinformatics*. 2014; 30(4):472–9. <https://doi.org/10.1093/bioinformatics/btt709> PMID: 24318998
30. Gu B, Sheng VS. A Robust Regularization Path Algorithm for v-Support Vector Classification. *IEEE Transactions on Neural Networks and Learning Systems*. 2016. <https://doi.org/10.1109/TNNLS.2016.2527796> PMID: 26929067
31. Gu B, Sun X, Sheng VS. Structural Minimax Probability Machine. *IEEE Transactions on Neural Networks and Learning Systems*. 2016. <https://doi.org/10.1109/TNNLS.2016.2544779> PMID: 27101618
32. Gu B, Sheng VS, Wang Z, Ho D, Osman S, Li S. Incremental learning for v-Support Vector Regression. *Neural Networks*. 2015; 67:140–50. <https://doi.org/10.1016/j.neunet.2015.03.013> PMID: 25933108
33. Zou Q, Wan S, Ju Y, Tang J, Zeng X. Pretata: predicting TATA binding proteins with novel features and dimensionality reduction strategy. *Bmc Systems Biology*. 2016; 10(4):114.
34. Xun W, Ying M, Minquan C. Finding Motifs in DNA Sequences Using Low-Dispersion Sequences. *Journal of Computational Biology*. 2014; 21(4):320–9. <https://doi.org/10.1089/cmb.2013.0054> PMID: 24597706
35. Liu B, Xu J, Lan X, Xu R, Zhou J, Wang X, et al. iDNA-Prot|dis: Identifying DNA-Binding Proteins by Incorporating Amino Acid Distance-Pairs and Reduced Alphabet Profile into the General Pseudo Amino Acid Composition. *PLoS ONE*. 2014; 9(9):e106691. <https://doi.org/10.1371/journal.pone.0106691> PMID: 25184541
36. Song T, Gong F, Liu X, Zhao Y, Zhang X. Spiking Neural P Systems with White Hole Neurons. *IEEE Transactions on Nanobioscience*. 2016; 15(7):666–73. <https://doi.org/10.1109/TNB.2016.2598879> PMID: 28029614
37. Song T, Zheng P, Wong MLD, Wang X. Design of Logic Gates Using Spiking Neural P Systems with Homogeneous Neurons and Astrocytes-like Control. *Information Sciences*. 2016; 372:380–91.
38. Xun W, Tao S, Faming G, Pan Z. On the Computational Power of Spiking Neural P Systems with Self-Organization. *Scientific Reports*. 2016; 6:27624. <https://doi.org/10.1038/srep27624> PMID: 27283843
39. Sogni P, Soupison T, Moreau R, Moine OL, Bacq Y, Hadengue A, et al. Homogenous spiking neural P systems with anti-spikes. *Neural Computing and Applications*. 2014; 24(7):1833–41.
40. Song T, Liu X, Zeng X. Asynchronous Spiking Neural P Systems with Anti-Spikes. *Neural Processing Letters*. 2015; 42(3):633–47.
41. Ma T, Wang Y, Tang M, Cao J, Tian Y, Al-Dhelaan A, et al. LED: A Fast Overlapping Communities Detection Algorithm Based on Structural Clustering. *Neurocomputing*. 2016; 207:488–500.