Paper 279-2013

Predictability by using social profile in Online P2P Lending Market

Zhangxi Lin, Texas Tech University; Siming Li, Southwestern University of Finance and Economics; Harshal Darade, Texas Tech University

ABSTRACT

Online peer-to-peer (P2P) lending markets enable individual consumers to borrow from, and lend money to, one another directly. We study the borrower-, loan- and group- related determinants of performance predictability in an online P2P lending market by conceptualizing financial and social strength to predict borrower rate and whether the loan would be timely paid. The result of our empirical study, conducted using a database of 9479 completed P2P transactions in calendar year 2007, provide support for the proposed conceptual model in this study. The results showed that combing financial files with social indicators can enhance the performance predictability in P2P lending market. Although social strength attributes do affect the borrower rate and status, their effects are very small in comparison to financial strength attributes. This paper concludes with specific suggestions to borrowers and lenders to increase their chances of funding to refunding completely in P2P lending market, and a discussion of future research opportunities.

INTRODUCTION

Person-to-person lending is a certain breed of financial transaction (primarily lending and borrowing, though other more complicated transactions can be facilitated) which occurs directly between individuals or "peers" without the intermediation of a traditional financial institution. The development of the market niche was further boosted by the global economical crisis in 2007-2010 when person-to-person lending platforms promised to provide credit at the time when banks and other traditional financial institutions were having fiscal difficulties. In 2005, there were $118 million of outstanding peer-to-peer loans. In 2006, there were $269 million, and, in 2007, a total of $647 million. The projected amount for 2010 is $5.8 billion.

One of the main advantages of person-to-person lending for borrowers has been better rates than traditional bank rates can offer. The advantages for lenders are higher returns than obtainable from a savings account or other investments. Person-to-person lending also attracts borrowers who, because of their past credit status or the lack of thereof, are unqualified for traditional bank loans. At the same time, because past behavior is frequently indicative of future performance and low credit scores correlate with high likelihood of defaulting, many person-to-person intermediaries have begun to decline borrowers whose credit scores are below a certain bound.

Early P2P lending was characterized by disintermediation and reliance on social networks but these features have started to disappear with the development of P2P lending market. Social network is increasingly important for lenders to find out good potential borrowers in the market, and for borrowers to get loan with less interest rate by reducing information asymmetric in the market.

So, in this paper, we study the borrower-, loan- and social- related determinants of performance predictability in an online P2P lending market by conceptualizing financial and social strength to predict whether the borrowers could be funded with lower interest, and the lenders would be timely paid. For this task, a conceptual framework is proposed, then model-based clustering method, discriminate analysis and confirmatory factor analysis was applied on real P2P lending data provided by Prosper.com.

BACKGROUND

Prosper marketplace’s database is available to researchers to deepen understanding of P2P lending market. Even though different P2P lending social networks may utilize different naming conventions, the underlying details are conceptually similar. Consequently, any conclusions derived from analyzing Prosper’s database may be generalized for the P2P lending industry.

Prosper is the largest P2P lending market in US based upon issued loan volume and revenue, it has issued 59,795 loans for $379,152,190. So, with the development of market size of P2P lending market, social network is increasingly important for lenders to find out good potential borrowers in the market, and for borrowers to get loan with less interest rate. Combing social network files with credit files can reduce information asymmetric in the market, enhance lenders’ predictability of risk evaluation of a money request and decrease the trading cost of loan transactions.

A unique feature of Prosper is its use of social networking through groups and friends. Anon-borrowing individual may set up a group on Prosper and become a group leader. The group leader is responsible for setting up the group web page, recruiting new borrowers into the group, coaching the borrower members to construct a Prosper listing, and monitoring the performance of the listings and loans within the group. The group leader does not have any legal responsibility. Rather, the group leader is supposed to foster a “community” environment within the group so that the group members feel social pressure to pay the loan on time. Group leaders can also provide an "endorsement" on a member's listing and bids by group leaders and group members are highlighted on the listing page. Since October 19, 2006, Prosper has posted star ratings (one to five) in order to measure how well groups perform against expected (Experian historical) default rates.

Starting February 12, 2007, Prosper members could begin to invite their offline friends to join the website. The inviting friend receives a reward when the new member funds ($25) or borrows her first loan ($50). Existing Prosper members can become friends as well if they know each other's email address but the monetary reward does not apply. Friends can also provide endorsements on each other's listings and a bid by a friend is highlighted on the listing page. Beginning February 23, 2008 lenders could begin including aspects such as friend endorsements and bids from friends as criteria in their listing searches.

Like other online networking engines, P2P lending market follows the fundamental properties of social networks, and behaves according to two phenomena: social influence and selection.

- *Social influence* takes place when a person diffuses its ideas to another person through interactions.
- *Selection phenomenon* is a reason why people tend to form communication and have relationship with persons who have similar interests with them.

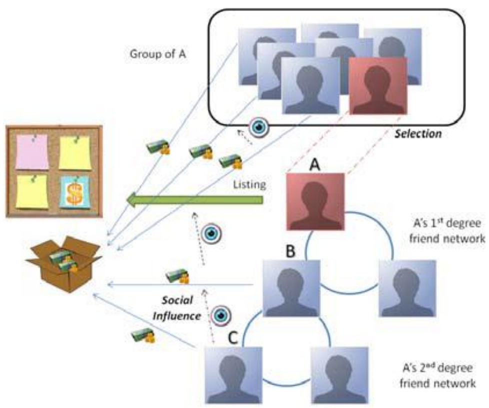


Figure 1. Two components of social network in Prosper

The function of two phenomena of social network in Prosper is shown in the above Figure 1.

- "Friend": it represents a one-to-one link from a member to other borrowers or lenders. This relationship between members is usually based on family, friendship or previous transaction history. It is made social influence and intends to motivate lenders to bid based on indirect trust.
- "Group": members are allowed to form communities with similar interest. Groups are formed by the selection phenomenon, where members are more likely to trust those who share some similarities with them.
- "Endorsement": members are allowed to give public feedback on previous transactions with other members, which may alter the impression between members in the market by social influence.

BORROWERS' CHARACTERISTICS IN P2P LENDING MARKET

Two type of borrower characteristics are considered in this project: social strength and financial strength.

- Financial strength. We considered three borrower characteristics that are more directly related to his or her financial strength---credit grade, debt to income ratio, and whether the person owns a house. While the first two are direct indicators of a borrower's creditworthiness, the third one, homeownership, is indicative of stability and a prior ability to access credit to obtain a mortgage.
- Social strength. We also considered three social network characteristics to evaluate borrowers' social strength---group rating, number of friends and whether the member is endorsed.

Trough Confirmatory Factor Analysis (CFA), we will conceptualize financial and social strength to predict performance in the P2P lending market both borrower and lender side respectively. For the performance indicators, we use Borrow Rate to evaluate borrowers' performance by analyzing whether and how financial strength and social strength factors effect interest rate borrowers can be funded fully; on the other hand, Borrower's Status is used to evaluate lenders' performance, and how financial and social factor impact on lenders' performance.

DATA PROCESSING

The Prosper dataset contains all the transaction and member data since its inception in November 2005. This is a considerable volume of information that encloses approximately (by December2008) 6 Million bids, 900,000 members, 4,000 groups and350,000 listings. In order to facilitate the analysis of the data, the dataset was filtered to contain all the loans created in calendar year 2007 and all the listings created in calendar year 2007, the bids created for these listings, the subset of members that created these listings and bids, and finally, the groups these members are affiliated with.

Once the data is converted from original XML file, we have 6 data sets available. In addition we would need information about different biddings against a listing which is available as a separated XML file. After data has been extracted we have 7 objects in all.

1. Category
2. Group
3. Listing
4. Loan
5. Marketplace
6. Member
7. Bid

For our Loan performance predictive model, we need to get the data that is related to borrowers' financial profile, social profile and loan profile. This information will be available from loan, listing, group and member objects. By the initial analysis of the data, the above observations are made regarding Data quality and data preprocessing. However, there are few fields (like "Borrower City", "Group Short Name", "Group State") are excluded for redundant or less important related to our analysis without losing information on the activity and interactions between members.

The filtering resulted inconsiderable reduction of the dataset: 9479 observations are available after selecting only borrower members. The variables finally used in our model are shown in the following table, where Status and Borrow Rate are our target variables.

Variables:
• Status (S): the status of borrower, which is classified into "GOOD" and "BAD" in this study.
• Borrower rate (BR): the rate borrower pays if the loan were to close at time.
• Debt to Income Ratio (DTI): the debt to income ratio of the borrower at the time the listing was created.
• Credit Grade (CG): Credit grade of the borrower at the time the listing was created. It goes from 7 (best) to 1 (worst).
• Is Borrower a Homeowner (IBH): specifies whether or not the member has is a verified homeowner at the time the listing was created.
• Group Rating (GR): quantitative evaluation of the performance of the group given by Prosper based on the transaction history. 5 (best) to 1(worst), and 0 for no grouping belong.
• Number of friends (NOF): number of
• Is Member Endorsement (IME): whether the borrower received endorsement.

Table 1. Input variables

And, the following table shows the summary of data description for each input variables included in our paper divided by group status. And then, the Table 3 represents the correlation between all the variables.

Is Group	Variable	Mean	Std. Dev.	N
0 N=5573	Debt-to-Income Ratio	0.3376563	1.0179519	5423
	Credit Grade	4.3457743	1.7759810	5573
	Is Borrower Homeowner	0.4826844	0.4997449	5573
1 N=3906	Debt-to-Income Ratio	0.5881534	1.7277741	3834
	Credit Grade	3.6267281	1.8544869	3906
	Is Borrower Homeowner	0.4237071	0.4942084	3906

	Group Rating	1.2800819	0.9509871	3906
	Number of Member Friends	2.7718894	4.8994957	3906
	Is Member Endorsed	0.0801947	0.2716292	3903

Table 2. Summary Statistic Result of Input Variables by Different Group Status

	DTI	CG	IBH	GR	NOF	IME
DTI	1.00					
CG	0.02 (0.06)	1.00				
IBH	0.03 (0.02)	0.33 (0.00)	1.00			
GR	0.07 (0.00)	-0.11 (0.00)	-0.06 (0.00)	1.00		
NOF	0.06 (0.00)	-0.11 (0.00)	-0.04 (0.00)	0.26 (0.00)	1.00	
IME	0.03 (0.01)	0.02 (0.13)	0.01 (0.39)	0.15 (0.00)	0.18 (0.00)	1.00

Table 3. Correlation of Input Variables

DATA EXPLORING

The following table shows the approx. records available based on the status out of 9479 observations. However, our interest is to predict whether the loan would be timely paid or not. So we will be replacing fully Paid status to "Good", indicating borrower is a good borrower and for the rest will be replacing to "Bad", indicating a bad borrower.

Status	Approx. Records
Charge-Off	2465
Defaulted	1153
Repurchased	13
Fully Paid	5848

Status	Approx. Records
Good	5848
Bad	3631

Table 4. Distribution of Borrower's Status in P2P lending market

Following table and chart represents the overall distribution of Target variable 'Status' with respect to group status (is in a group/ is not in any group). And it is obviously to see that there are more borrowers belong to at least one group in "GOOD" Status.

	GOOD	BAD	TOTAL
ISGROUP	2211	1678	3889
ISNOTGROUP	3637	1953	5590
TOTAL	5848	3631	9479

Table 5. Distribution of Borrower's Status in P2P lending market

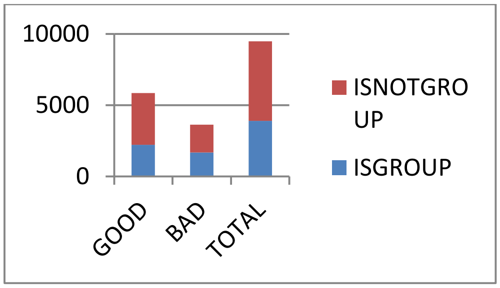


Figure 2. Distribution of Borrower's Status in P2P lending market

Then the detail of statistic result of input variables in different Borrower Status category is shown in the following Table 5. In the table, it is reasonable to see that good borrowers have higher average credit grade and average group rating but lower debt to income ration than bad borrowers. However, this table also shows that bad borrowers have more chance to get member's endorsement and have more friends, what might be happen when borrowers want to use social information to compensate lack financial credit to get fund which beyond their solvency.

Status	Variable	Mean	Std. Dev.	N
BAD N=3631	Debt-to-income Ratio	0.5457773	1.6160739	3538
	Credit Grade	4.4467089	1.7299372	3631
	Is Borrower Homeowner	0.4731479	0.4993472	3631
	Group Rating	1.1066746	0.5535731	1678
	Number of Member Friends	2.9815256	5.5206247	1678
	Is Member Endorsed	0.0990453	0.2988120	1676
GOOD N=5848	Debt-to-income Ratio	0.3768369	1.1755415	5719
	Credit Grade	5.4329685	1.8094503	5848
	Is Borrower Homeowner	0.4492134	0.4974566	5848
	Group Rating	1.4215287	1.1445593	2211
	Number Of Member Friends	2.6140036	4.3685225	2228
	Is Member Endorsed	0.0665468	0.2492915	2224

Table 6. Summary Statistic Result of Input Variables by Different Borrower Status

CONCEPTUAL MODEL

Based on the previous discussion, the conceptual framework of this project is shown in the following Figure 3:

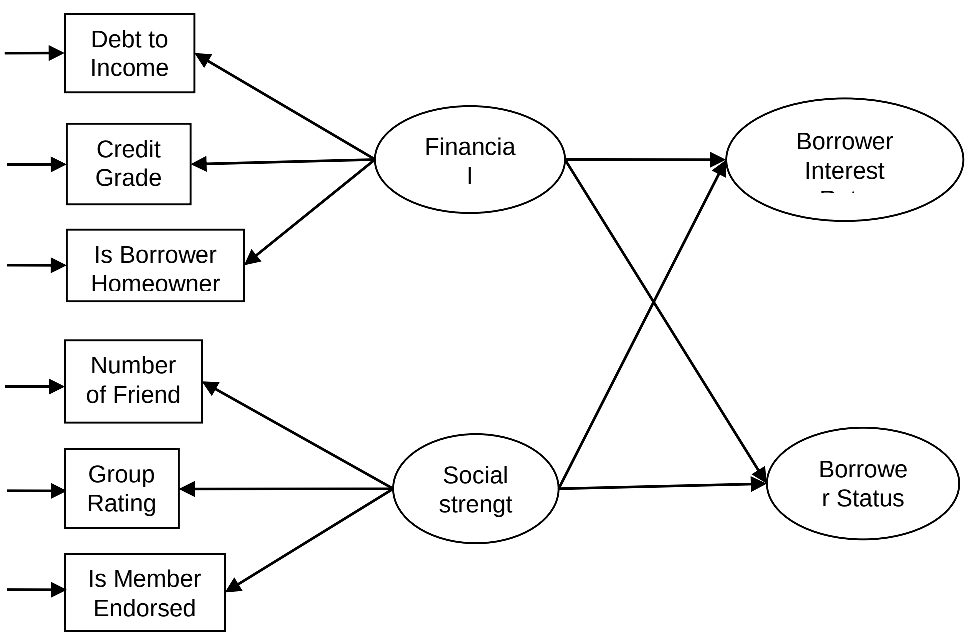


Figure 3. Conceptual Framework in P2P Lending Market

As we conceptualize in the above model and have developed in the previous discussion of important relationship, the conceptual framework indicates that borrower characteristics potentially have effects on loan performance. The following three hypotheses summarize this discussion:

H1: Combining financial factors with social factors can enhance the ability to predict performance in P2P lending market

H2: Social strength is more important than financial strength when predict lenders' performance.

H3: Social strength is more important than financial strength when predict borrowers' performance.

H4: Combining financial strength and social strength factors can distinguish borrowers.

METHODOLOGY

BORROWER STATUS PREDICT

Canonical Discriminant Analysis (CDA) is used to define Financial Strength and Social Strength variables by using linear combination of different independent variables. And the result is show in the following Table 7:

Total Canonical Structure		
Variable	Label	Can1
Debt-to-Income Ratio	Debt-to-Income Ratio	-0.188711
Credit Grade	Credit Grade	0.895953
Is Borrower Homeowner	Is Borrower Homeowner	-0.087968

Table 7. CDA result for financial strength in P2P Lending Market

Then, we define the financial strength by combing the three financial-related determinants, and the relationship is shown in the following Equation 1:

$$Financial\ Strength = -0.19\ DIT + 0.90CG - 0.09IBH \qquad (1)$$

Similarly, we found out Equation 2 to describe Social Strength by CDA method whose result is shown in the table 8.

Total Canonical Structure		
Variable	Label	Can1
Group Rating	Group Rating	0.932775
Number Of Member Friends	Number Of Member Friends	-0.218556
Is Member Endorsed	Is Member Endorsed	-0.342570

Table 8. CDA result for social strength in P2P Lending Market

$$Social\ Strength = 0.93\ GR - 0.22NOF - 0.34IME \qquad (2)$$

Based on the financial and social strength data, we exam the relationship between Borrower's Status and Borrower's characteristics by logistic regression analysis. And the result is shown in Table 9. The results are significant, and then we say that financial and social factors have significant effects on predicting Borrowers' Status. What's more, in this case, financial and social factors almost have same impact on target variable.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3456	0.0353	95.7070	<.0001
financial	1	-0.6103	0.0358	291.4018	<.0001
social	1	-0.4454	0.0455	95.7741	<.0001

Table 9. The relationship between Borrowers' Status and Borrowers' Characteristics

Also, for comparing with the Borrower Rate case, we also do the linear regression for Borrowers' Status variable.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr. > |t|
Intercept	Intercept	1	0.57109	0.00756	75.56	<.0001
financial	financial	1	0.13279	0.00725	18.32	<.0001
social	social	1	0.07803	0.00753	10.37	<.0001

Table 10. The relationship between Borrower Status and Borrowers' Characteristics

Figure 4. Scatter Distribution of Residuals by Regression for Status

Then, we express the relationship by Equation 3.

$$Borrower\ Status = 0.57 + 0.13\ Financial\ Strength + 0.08\ Social\ Strength \quad (3)$$

This equation shows us that higher financial and social strength can decrease the chances of default in Prosper market. To predict the borrowers' status, the financial strength is more important than social strength.

BORROWER RATE PREDICT

For borrower rate, we propose simple linear regression model to see how financial and social strength factors affect the degree of interest rate. And the regression result is shown in the Table 9 and the Figure 5 represents the histogram of residual. From these results, the regression models perform very well, and all the parameters are significant in this case. And R-square=0.5 means that lenders using financial and social strength to evaluate half of debt paying ability of borrowers in P2P Lending market, and then decide the lend rate to a money request. Comparing to financial characteristics, social factors has less impact on lenders' decision making, however, higher financial and social strength produce lower borrowing interest respectively.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr. > |t|
Intercept	1	0.18306	0.00071603	255.66	<.0001
financial	1	-0.04216	0.00068676	-61.39	<.0001
social	1	-0.00801	0.00071320	-11.23	<.0001

Table 11. The relationship between Borrower Rate and Borrowers' Characteristics

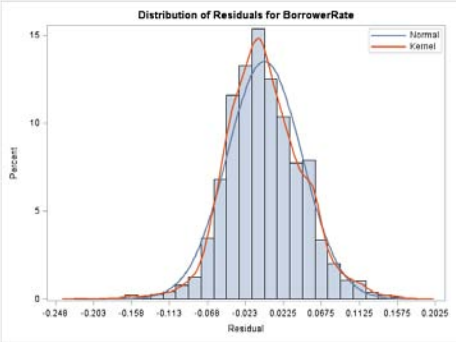


Figure 5. Distribution of Residuals for Borrower Rate

Then, we express the relationship by Equation 4.

$$BorrowerRate = 0.57 - 0.04\ Financial\ Strength - 0.01\ Social\ Strength \quad (4)$$

BORROWER DISCRIMINATION

Figure 6 shows the scatter distribution of social strength variable by financial strength variable, which are created in the 5.1 section. We cannot find any clustering result from this figure. And we use the discrimination analysis method to show whether or not we can use only financial strength and social strength variable to predict the status of potential borrowers in the P2P market. With 57.51% error rate, Table 12 tells that combing these two characteristics is not enough to distinguish borrowers.

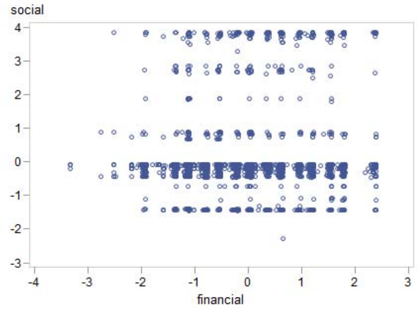


Figure 6. Scatter of Financial Strength by Social Strength

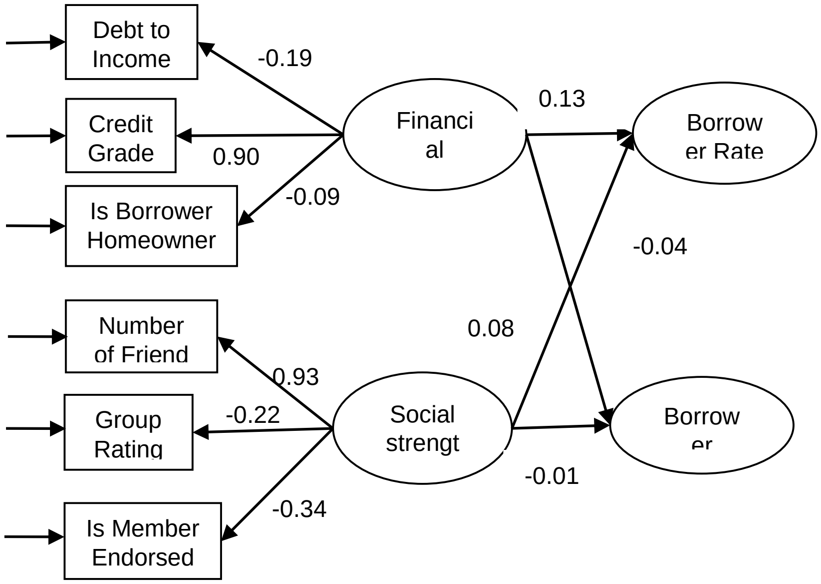
Number of Observations and Percent Classified into Status				
From Status	Charge-off	Defaulted	Paid	Total
Charge-off	583 54.23	167 15.53	325 30.23	1075 100.00
Defaulted	298 52.37	87 15.29	184 32.34	569 100.00
Paid	635 29.04	285 13.03	1267 57.93	2187 100.00
Total	1516 39.57	539 14.07	1776 46.36	3831 100.00

Error Count Estimates for Status				
	Charge-off	Defaulted	Paid	Total
Rate	0.4577	0.8471	0.4207	0.5751

Table 12. Discrimination Analysis of Borrowers' Status

EMPIRICAL RESULT

Based on results in sections 5.1 and 5.2, we update our conceptual model in the Figure 5 as following.



Then we conclude that:

H1: Combining financial factors with social factors can enhance the ability to predict performance in P2P lending market.

We accept H1, for both financial and social variables are significant for explaining borrowers' status and borrower rate, respectively.

H2: Social strength is more important than financial strength when predict lenders' performance.

We reject H2. In the conceptual model, we can say that financial strength (0.13) is more important than social strength (0.08) when predict borrowers' interest rate.

H3: Social strength is more important than financial strength when predict borrowers' performance.

We reject H3. In the conceptual model, we can say that financial strength (-0.04) is more important than social strength (-0.01) when predict borrowers' status.

H4: Combining financial strength and social strength factors can distinguish borrowers.

We reject H4. In section 5.3, by using discrimination analysis method, borrowers cannot be distinguished only by financial and social strength variables.

LIMITATION AND FUTURE WORK

Our research contributes to the growing literature on performance predictability by combining financial and social files in the P2P lending market. This work uses three variables to represent financial strength and social strength factors in P2P lending market respectively. There are a lot of related variables uncovered in this study, which might reduce the accuracy of the proposed conceptual model. Unfortunately, the parameter estimates for financial strength variable and social strength variable cannot to be compared, though both models show strong depends on independent variable (borrowers' status and borrower rate). Also, moral hazard and adverse selecting are there in the process of grouping selecting and the cost of building social network in P2P market is lower than in real world. So, the situation is complicated when analyzing the effect of social interaction on lending and borrowing behavior in P2P market, other existing social network such as Facebook, Twitter or MySpace should be included to define the social strength variable. Obviously, the solution of this study is too simple and the dataset is also weak.

Recent economic events worldwide have raised the research topic on whether P2P lending market may be a beneficial financial service and might positively impact on individuals, traditional financial institutes and financial markets. Several future works are listed:

- Moral hazard and adverse selecting in P2P lending market, and how they impact on the behavior of borrower and lender.
- Hidden Social Interactions: the relative amount of loan created by second degree friends, group leaders and other group members, are hidden social features that need to be considered in the future.
- Risk management is essential for lenders in the P2P market as in traditional financial market.
- Performance evaluation analysis of P2P lending market.

REFERENCE

S. Freedman, G.Z.Jin. "Do social networks solve information problems for Peer-to-Peer lending?" Evidence from Prosper.Com-papers.ssrn.com.

Hampshire, Robert (2008). "Group reputation effect in Peer-to-Peer lending markets: an empirical analysis from a Principle-Agent Perspective." working paper.

Mendelson, Haim (2006). "Prosper.com: A People-to-People lending marketplace." working paper.

Heerero-Lopez (2009). "Social interaction in P2P lending." working paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Siming Li
Enterprise: Southwestern University of Finance and Economics
Address: 3251, 62nd Street
City, State ZIP: Lubbock, TX79413
Work Phone: 806-787-6759
E-mail: lsm_lsm_lsm@126.com