Follow the Profit or the Herd? Exploring Social Effects in Peer-to-Peer Lending

Dawei Shen, Coco Krumme, Andrew Lippman
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA, 02139
Email: {dawei, coco, lip}@media.mit.edu

Abstract—This paper examines the impacts of social factors on lenders’ decision-making in online peer-to-peer (P2P) lending. Data collected from a major U.S. online loan marketplace, Prosper.com, have been analyzed. We propose a model based on preferential attachment and fragmentation to model the bidding behavior of lenders. Our data analysis presents strong empirical evidence that there were significant herding effects when lenders made their investment decisions on loan listings. The distribution of the number of bids put on loan listings exhibits a power law with an exponential cutoff, which matches what the model predicts. The paper concludes that lenders on Prosper did not make rational investment decisions based on risk and returns, but followed the herd.

I. INTRODUCTION

A. Overview of Peer-to-Peer Lending and Prosper.com

In a Peer-to-Peer (P2P) lending system, people bank on each other. During the recent recession when the availability of credit was tightly controlled, more and more people discovered the benefits and flexibility of P2P lending, though P2P lending has existed in various formats for thousands of years. When an individual is in need of money, instead of seeking loans or applying for credit from a centralized entity such as a bank, he/she can directly borrow money from his/her peers who are willing to lend. Today, P2P lending has emerged as an interesting area both in academia and in financial industry.

Prosper.com (abbr. Prosper)[1] is such an online P2P lending marketplace where people can invest in each other in a way that is both socially and financially rewarding. Borrowers and lenders interact mainly through Prosper’s online auction platform. An individual solicits loans by posting a listing on Prosper with the amount he/she requests, the maximum interest rate he/she is willing to accept, and a statement of the purpose of the loan. Investors, as lenders, can access borrowers’ credit profiles and bid on posted listings with the amount they lend, and the minimum interest rate they are willing to offer. When the auction expires, a listing gets materialized into a loan if and only if the amount requested gets fully covered by one or multiple lenders, and Prosper selects the bids with the lowest interest rates. Prosper, as the intermediary, collects payment from borrowers and distributes it among lenders.

The activity of lending on Prosper differs from lending in the traditional banking industry in several important ways. The need for collateral is removed and the system is based on mutual trust. Prosper also provides avenues for creating social networks, joining interest-based groups (tied by geography, common interest, or common loan purpose), and collecting endorsements from friends and group leaders. Individual lenders generally lack the complex risk assessment models that banks utilize for their operations. Therefore, lenders judge investment opportunities also by these ‘social’ factors. Another important social factor that impacted Prosper was that lenders mutually influenced each other’s decision-making when they decided which listings to invest. Modeling and understanding the effects and implications of social influence is the primary goal of this paper.

Prosper.com made its data publicly available for analysis and research to encourage better understanding of this emerging financial service[2]. The data have been anonymized to preserve the privacy of users. The dataset captures all activities which occurred on Prosper from 2006 to 2008 including all information on bids, listings, social links, grouping, etc.

B. Social impacts on lenders’ decision making

In this paper, we are focused upon modeling and understanding lenders’ behavior on Prosper and how their decisions are influenced by social factors.

Investment on Prosper was rewarding but highly risky. The average annual percentage rate (APR) for borrowers who had a FICO credit score of larger than 640 was 14.5%. Borrowers with poor credit grading who had a FICO score of less than 640 offered an even higher return: 23.7% APR. However, investment on Prosper also presented a significant risk. Fig.1 summarizes the performance of all loans on Prosper used in the study. Nearly 20% of loans were charged-off and around 5% of all loans were in the status of ‘Active late’, which implied a potential default risk.

It is widely believed that a rational investment decision should be an exclusive function of financial risk and expected returns. When lenders on Prosper decided which listings to invest, they were expected to carefully examine certain financial features of the lending opportunity such as borrowers’ credit ratings, debt-to-income ratio, whether the borrower is a home-owner and offered interest rates to assess risk and return before making a bid.

However, there were prominent information asymmetry problems existent on Prosper that imposed great challenges to lenders[3]. Adverse selection pushed the interest rate higher.
and attracted more and more risky borrowers with low credit ratings to stay around the marketplace seeking loans. Lenders, with insufficient information about borrowers, encountered great difficulty assessing the potential risk. Therefore, they used other factors, the so-called ‘social features’, to make a more informed decision. These ‘social features’ include group ratings, endorsement from friends and group leaders, etc.

Another important observation is that many lenders made their investment decisions by simply following others’ activities. Lenders tended to put their bids on listings that had already attracted a fair amount of bids. They did so possibly because they might believe that others had obtained extra credible information about a borrower, or they might prefer not to waste their time in listings with few bids which were unlikely to turn into loans.

Such behaviors are usually referred to as herding effects in social science and financial market studies. A key goal of this paper is to study techniques for modeling herding effects, and examine real-world data collected from Prosper, and validate whether herding effects exist among Prosper lenders by comparing what the model predicts with observations from data.

C. Outline

The remainder of this paper is organized as follows. In the next section, we briefly review prior literature on the modeling of herding effects and related research on P2P lending in general. Section III discusses the impacts of social factors on P2P lending through statistical analysis. The following two sections focus on the analysis of herding effects in P2P lending. A herding model based on preferential attachment and fragmentation is introduced in Section IV. The following section verifies the model through data analysis and presents the empirical evidence of herding effects. The last section concludes the paper with implications and directions for future study.

II. RELATED WORK

P2P lending has emerged as a popular research topic. A number of research papers conducted analysis based on the dataset Prosper made available. [4] found that the intermediary, which referred to the electronic marketplace Prosper.com, created value by reducing information asymmetries between borrowers and lenders. The recommendation of a loan listing and the intermediary’s bid had a strong impact on the resulting credit spread. [5] treated investment on Prosper as a new type of fixed-income fund: a P2P lending portfolio, which created an educational experience for students to run an investment fund. [6] analyzed risk and return of investment opportunities on Prosper from a lender’s perspective, and recommended rules that lenders should follow to make a profit.

Herding behavior describes how individuals in a group act together without planned direction. Herding behavior exists in various human activities such as stock market bubbles and crashes, street demonstrations, sporting events, opinion forming, etc. Some classic papers that studied the theory and application of herding behavior include [7][8][9].

Herding behavior among lenders on Prosper was analyzed in [10]. The authors found that a 1% increase in the number of bids increased the likelihood of an additional bid by 15%. Another surprising result was that there existed a positive association between herding in the loan auction and its subsequent performance. Our paper will develop a similar conclusion using a different modeling technique by analyzing the distribution of bids among listings.

There are a significant number of papers that study herding effects as well as online auction from various perspectives. Here, we introduce a few of them which are most relevant and helpful to our study.

[11] represents a wide class of research work that models the arrival process of bids. In this paper, authors modeled bid arrivals using non-homogeneous Poisson processes and illustrated its validity by fitting and interpreting data from eBay.com.

Another line of research treats decision makers as individuals in a networking environment and models their herding behaviors in the form of group forming and fragmentation. Herding is mostly observed and studied in the context of financial markets. [12] thoroughly examined empirical and theoretical evidence of herding in a financial market. It revealed the fact that the distribution of stock market returns had fat tails observed and market participants tended to imitate each other. They considered the network of agents to be a diluted regular lattice, and showed the distribution of cluster sizes could lead to a power-law distribution of returns. [13], [14] and [15] proposed and further extended a so-called E.Z. model, which introduced a kinetic model in which a group of agents can either coagulate or fragment at each time step. They gave a cluster size distribution characterized by a power law with an exponent of $-\frac{5}{2}$.

A herding model based upon preferential attachment and fragmentation is proposed in [16] and [17]. This model best captures and characterizes the dynamics of the bidding process on the Prosper platform. In this paper, we analyze this model in detail, and then examine the distribution of the number of bids on Prosper to validate the model.

III. IMPACTS OF SOCIAL FACTORS IN P2P LENDING

Social factors have significant impacts on a P2P lending marketplace. For example, information and beliefs propagate
among members and influence their decision making. Members perform self-selection to form groups or social links and monitor each other’s behavior. When we analyze social effects on Prosper.com, we focus on the following components

- **Friendship** - Members could establish friendship among themselves by adding/approving friends.
- **Grouping** - Members around an affinity could create a group. A group was managed by ‘Group Leaders’. Current group members could decide if a new member would be accepted into the group. The reputation of a group is measured by Prosper based on the loan performance of its members. Each group is rated on a 1 ∼ 5 scale, or labeled as non-rated.
- **Endorsement** - Borrowers can have endorsement from their friends, group leaders or other members.

The Prosper dataset contains all transaction and membership data from its opening in November 2005 to December 2008, which includes approximately 900,000+ members, 4,000+ groups, 350,000+ listings, and 6 million bids[18]. In this section, we analyze the impacts of social factors to the likelihood of loan conversion and loan performance. Here, loan conversion means a borrower’s listing gets materialized into a loan after attracting enough bids.

A. Influence of Credit Profiles

Before analyzing the impacts of social factors, we should first acknowledge that financial features keep playing a significant role on a P2P lending marketplace. Lenders evaluate the implied risk of a loan listing based on both the borrower’s financial features and social features.

Lenders considering a bid on a borrower’s loan listing have access to the borrower’s credit grade, along with summary credit data from the prospective borrower’s Experian credit history, including number of current delinquencies, amount currently delinquent, delinquencies in the past 7 years, and some other information. Prosper maps a user’s Experian credit core to a credit grade. Possible values include AA, A, B, C, D, E, HR (High Risk) and N/A(not available).

Fig.2 reveals the importance of the credit grade. The higher the grade was, the more likely a borrower was able to get the credit request fully funded. It was a remarkable fact that borrowers with extremely poor credit grading (E and HR) could still acquire loans with relatively high likelihood. These borrowers had a FICO credit score of lower than 600. It would be extremely difficult for these risky borrowers to acquire credit in the traditional way.

B. Influence of Group Membership

A member’s group affiliation was influential in getting his/her credit request fully funded, and possibly the interest rate with which this loan was eventually materialized. On Prosper.com, groups were rated from 1 star to 5 stars depending on the past performance of their members. Groups that had not been in the marketplace for enough time to be evaluated were represented by the ‘Not yet rated’ label. Groups were managed by ‘Group Leaders’ who brought borrowers to Prosper, maintained the group’s presence on the site, and collected and/or shared ‘Group Rewards’. Some ‘Group Leaders’ charged a ‘Group Leader Reward Rate’ from all loans solicited by group members.

As shown in Fig.3, as the borrower moves to a higher-rated group, the success rate of his/her listing getting fully funded increases. The column for the label ‘Has a Group’ refers to all users who had group affiliation, whose groups could be either rated or non-rated. It’s also evident that the chance of getting a loan in the marketplace for borrowers without group affiliation was very slim, merely 6%, while the possibility was elevated to nearly 18% if the borrower was affiliated to a 5-star group.

Group affiliation also significantly influenced the loan performance in the marketplace. As shown in Fig.1, the status of a loan could be one of the following: Current, Paid, Late or Charged-off (default). A loan was considered healthy only if it was current or paid off.

Within a group, there existed peer pressure among its members to build and maintain the group’s reputation and thus the rating of the group. This fact favorably improved the performance of loans acquired by members in higher-rated groups. Fig. 4 displays loan performance for groups with
different ratings by showing only the late/default rates. The percentage of late or default loans considerably decreased as the rating of the group improved. For groups rated with ‘1-star’, the average late/default rate reached one third, whereas this rate was merely less than 2% for a ‘5-star’ group. On Prosper, groups rated with ‘1-star’ often had open membership and anybody could join. These groups were usually largely populated, less managed and poorly organized, where the social pressure was not effective in helping the loans perform better. This is reflected by the observation that borrowers in the ‘1-star’ group and borrowers who were not affiliated with any group performed similarly [18].

C. Influence of Friendship and Endorsement

Making friends on Prosper and maintaining a friend network helped getting a loan request fulfilled. Among all loan listings that were fully funded and materialized into loans, 43% of solicitors had made at least one friend on Prosper, and 17.8% of these listings had received bids from the borrowers’ friends. For loan listings that failed to become loans, the two ratios were only 30% and 3.9% respectively [18].

Another important social factor is endorsement. On Prosper, members could endorse each other by leaving favorable comments on each other’s profile pages. They served as reference letters for lenders who reviewed their profiles. Instead of doing textual analysis of endorsement, we treat the effect of endorsement as a binary variable: with or without endorsement. Listings with endorsement had a success rate of 21.2%, while listings without endorsement only had a chance of 11.5% of turning into a loan. It is very evident that having endorsement almost doubled the possibility of getting a loan. Therefore, endorsement was very influential during the loan solicitation process even if you don’t read it.

In the next two sections, we will model and analyze the herding effects in greater details.

IV. MODELING HERDING EFFECTS

A. A herding model with preferential attachment and fragmentation

Let’s first consider a dynamic system with groups of agents. At each time step, one of two events can occur. With probability \( p \), an agent is added to the system and joins a group of size \( k \), with a rate that is proportional to \( k \). With probability \( q = 1 - p \), a group is fragmented into individual agents. Let \( n_k(t) \) denote the number of groups of size \( k \) at time \( t \), it evolves as:

\[
\frac{dn_k(t)}{dt} = p \frac{M(t)}{N(t)}[(k-1)n_k(t) - kn_k(t)] - q \frac{n_k(t)}{N(t)}
\]

for \( k \geq 2 \). The number of agents who do not belong to any groups (or they can be treated as individuals who are in groups of size 1) evolves as

\[
\frac{dn_1(t)}{dt} = -p \frac{n_1(t)}{M(t)} + q \frac{N(t)}{N(t)} \sum_{k=2}^{\infty} k n_k(t).
\]

In these equations,

\[
N(t) = \sum_{k=1}^{\infty} n_k(t), \quad M(t) = \sum_{k=1}^{\infty} kn_k(t).
\]

\( N(t) \) represents the total number of groups formulated in the system, and \( M(t) \) represents the total number of agents who have joined in the system. In the right-hand side of Eq.(1), the first term describes the event that a new agent joins an existing group, and the second term models the fragmentation of a group of size \( k \). It also captures the phenomenon of preferential attachment. In the right-hand side of Eq.(2), the first term is the result of the arrival of a new agent, and the second summation term is the creation of individual agents by the fragmentation of other groups.

Eq.(1) and Eq.(2) are constructed in a seemingly rather complicated form. However, after we take the summation \( \sum_{k=1}^{\infty} \frac{dn_k(t)}{dt} \) and \( \sum_{k=1}^{\infty} \frac{dn_1(t)}{dt} \), the result gets greatly simplified. After cancelling out a number of summation terms, we are left with an equation in a succinct and tractable form:

\[
\frac{dN(t)}{dt} = q \cdot \left( \frac{M(t)}{N(t)} - 1 \right), \quad \frac{dM(t)}{dt} = p.
\]

The simple form of Eq.(4) directly suggests the solutions of \( N(t) \) and \( M(t) \), which both turn out to be linear in time \( t \):

\[
M(t) = pt, \quad N(t) = at, \quad \text{where} \quad a = \frac{-q + \sqrt{q^2 + 4pq}}{2}.
\]

Further more, Eq.(4) also suggests that the solution for \( n_k(t) \) is also linear in time. We can write \( n_k(t) \) as \( n_k(t) = c_k t \), and obtain the recursive relationship for \( c_k \):

\[
c_k = [(k-1)c_{k-1} - kc_k] - \frac{q}{\alpha} c_k,
\]

as well as the initial condition

\[
c_1 = -c_1 + \frac{q}{\alpha} (p - c_1).
\]

A cleaner form can be derived from Eq.(5) and Eq.(6):

\[
\frac{c_k}{c_{k-1}} = \frac{k-1}{k + 1 + q/\alpha}, \quad c_1 = \frac{pq}{\alpha} \frac{1}{2 + q/\alpha}.
\]
Then, we can obtain the solution for $c_k$ iteratively:

$$c_k = \frac{pq}{\alpha \beta} \prod_{l=1}^{k-1} \frac{l}{p + \beta},$$

(8)

where $\beta = 2 + q/\alpha$. When $k$ is sufficiently large, $c_k$ can be well approximated by a power law

$$c_k \approx k^{-\beta}.$$  

(9)

This approximation becomes obvious when $\beta = 1$. In this case we can compute $c_k$ in its exact form: $c_k = \frac{2q}{\gamma} k^{-1}$.

This result successfully establishes the hypothesis that the distribution of group sizes follows a power law, which also matches our empirical observations in a financial market where herding effects play an important role in investors decision-making process.

B. Discussion

The derivation above mainly follows the idea presented in [16]. In our paper, we have carefully validated each step and applied the model in the context of bidding in a P2P lending environment.

A notable distinction is that in [16], the value of $q$ is fixed to be $q = 1 - p$, which can be relaxed. Instead of interpreting $p$ and $q$ as probabilities of two alternative events, we can treat them as constant rates for group joining and group fragmentation without affecting other derivation. We note that in this model $\beta \geq 2$.

This model emulates the online bidding process on Prosper very well. It captures the fact that lenders tend to bid on a listing with a likelihood that is in proportion to the number of bids the listing has already attracted. More importantly, each listing on Prosper expires after a certain period of time, which further implies that no group can last for infinitely long.

Solving a differential equation and checking solutions at a particular stopping time in an ever-changing and dynamic network environment can make the problem intractable. This model avoids this problem by assuming a group fragments at a certain rate with certain probabilities. Though this approximation is valid, it does not reflect the fact that a listing with more bids typically implies that it has existed for a longer time, and thus more likely it will disappear/fragment soon. This problem is resolved in the generalized model we introduce below.

A caveat of this model is that all independent agents are assumed to be in their single-member groups. This means that the model does not distinguish lenders who have not bid on any loan listings, and lenders who become first bidders of listings.

C. Generalization

The model can be generalized by allowing incoming agents to join groups of size $k$ with a rate $A_k$ and fragmenting group of size $k$ with a rate $B_k$. Rate equations Eq.(1) and Eq.(2) can be rewritten as

$$\frac{dn_k(t)}{dt} = p \frac{A_k}{A(t)} n_{k-1}(t) - A_k n_k(t) - q B_k \frac{n_k(t)}{B(t)},$$

(10)

and

$$\frac{dn_k(t)}{dt} = -pA_k \frac{n_k(t)}{A(t)} + q B_k \sum_{k=2}^{\infty} k B_k n_k(t),$$

(11)

where

$$A(t) = \sum_{k=1}^{\infty} A_k n_k(t), B(t) = \sum_{k=1}^{\infty} B_k n_k(t).$$

(12)

While Eq.(10) and Eq.(11) cannot be solved for general $A_k$ and $B_k$. There are a few interesting special cases for different choices of $A_k$ and $B_k$. We are mostly interested in the model when $A_k = B_k = k$. In this case, the model assumes that a group fragments with a rate that is also proportional to the group size, which is exactly what we wish to incorporate in the model. Following a similar process of calculation, we find the distribution of group size follows

$$c_k \approx \prod_{l=1}^{k-1} \frac{l}{1 + p + 1},$$

(13)

In this generalized model, the distribution of group sizes exhibits a power law with an exponential cut-off for large $k$. The base $\gamma$ satisfies $0 < \gamma < 1$.

So far we have developed a generalized herding model based on preferential attachment and fragmentation, which is suitable for capturing the bidding dynamics on Prosper. The complex time-variant dynamics of lenders’ behaviors are captured and reflected in a simple format: the distribution of group sizes. In the context of Prosper, it is the distribution of the number of bids on all loan listings that we are interested in calculating and comparing with what the model predicts.

V. DATA ANALYSIS

We have analyzed data that records all 6 million bids made by lenders from 2006 to 2008. We are mainly interested in the distribution of bids that each listing attracts. The data have shown that each listing can attract as many as 1020 bids. Out of the total 342341 listings on Prosper.com, 160082 listings, nearly half of them, never attract any bids.

We first look at the empirical distribution of the number of bids in the time domain. We only use bids that were made in Jan. 2008 and Feb. 2008, whose corresponding listings had a bidding duration of 10 days. We plot the number of bids made $k$ days after the creation of the listings, where $k$ ranges from 1 to 10, in Fig.5. Though the time domain distribution is not directly related to the herding model we present above, it helps us understand the mentality of lenders participating in the bidding process. The curve matches observations in other online bidding system, and the theoretical findings in [11]. The start of the bidding typically sees an unusual amount of early bidding, which is followed by a period of low activity. The auction end typically experiences a huge amount of last-minute bidding. We hypothesize that the preferential attachment process typically happens after the low-activity period, when lenders try to follow listings with more bids.
Next, we calculate the empirical distribution of the number of bids among loan listings, i.e., the distribution of group sizes as previous called in the model development. Fig.6 displays the histogram in a log-log scale.

Fig.6 clearly shows that \( k = 100 \) is a turning point for the group size distribution. When \( k < 100 \), the group size distribution curve looks almost perfectly linear in the log-log scale, which in turn implies that the distribution follows a power-law as the model predicts. When \( k > 100 \), the curve decreases exponentially with \( k \). This is more evidently shown in Fig.7, which plots the same curve in the linear-log scale. The segment for \( k > 100 \) shows strong linearity, which suggests that the distribution is decaying exponentially with \( k \) when \( k \) is sufficiently large.

We next approximate the linearity by using linear regression. We zoom into the two regions on the left-hand side and the right-hand side of \( k = 100 \), and perform linear regression in the log-log domain and the linear-log domain respectively.

The actual and fitted distribution curves are shown in Fig.8 and Fig.9. Evidently, both segments are well approximated by linear functions. The only exception is that when \( k \) gets close to 1000, a significant amount of fluctuations can be observed around the fitted linear line.

From linear regression, we obtain that for the power law distribution when \( k < 100 \), the power \( \beta \approx 1.14 \). The other segment for \( k > 100 \) is exponentially decaying with \( e^{-\eta k} \), where \( \eta = 0.0065 \).

For the purpose of comparison, Fig.10 and Fig.11 display the distribution of bit counts among listings that eventually turned into loans. We distinguish two cases: the blue curve is for loans that were current or paid, while the green curve is for loans that were late, delinquent, or charged off. They represent loans that exhibited exactly the opposite performance. As expected, power law distribution is no longer visible because these listings attracted enough bids to become loans. An
interesting observation is that lenders’ behaviors do not vary with risks.

To summarize what we have discovered from our data analysis. There existed strong empirical evidence for herding effects among lenders on Prosper when they bid on loan listings. We have seen an evident match between what the model predicts and what the empirical data tell us. Eq.(13) suggests that when $k$ is small, the distribution follows a power-law, and when $k$ becomes sufficiently large, the exponential term dominates, and it decays exponentially. This is clearly visible in our data analysis. Thus, we have reasons to believe that lenders on Prosper were strongly affected by herding effects, and they did not always rationally make their financial decisions.

VI. CONCLUSION

In this paper, we first examined the impacts of various social factors on Prosper, including group affiliation, friendship, and endorsement. We found that while financial features, which mainly refer to credit profiles of borrowers, played a significant role in lenders’ selection process, these social factors also exhibited significant influence to lenders’ investment decisions.

In the second half of the paper, we focused on the analysis of herding effects among lenders by proposing a herding model and fitting Prosper’s bidding data to validate the model. The contribution of this study is two-fold. We first proposed a herding model that was built upon preferential attachment and group fragmentation. The model best characterizes the properties and dynamics in the bidding process of Prosper. Second, we analyzed real-world data from Prosper.com, and verified that the group size distribution matched what the model would predict very well.

In conclusion, herding effects did exist in lenders’ decision-making process. Lenders on Prosper did not always make financial decisions rationally, but simply followed others’ actions. Future lenders on a P2P lending marketplace should be reminded that other lenders’ actions are not necessarily made because of extra credible information, but simply herding effects.

ACKNOWLEDGMENT

The research presented in this paper has been supported in part by the Center for Future Banking (CFB) at the MIT Media Lab.

REFERENCES