ORIGINAL ARTICLE

A decision tree model for herd behavior and empirical evidence from the online P2P lending market

Binjie Luo · Zhangxi Lin

Abstract Herd behavior has been studied in a wide range of areas, such as fashion, online purchasing, and stock trading. However, to date, little attention has been paid to the herd behavior in online Peer-to-Peer lending market. With a decision tree, we model the formation of herding when decision makers with heterogeneous preferences are facing costly information acquiring and analyzing. Data from Prosper.com provides us with a good opportunity to explore empirical evidences for herd behavior. When herd behavior arises, individuals follow the behaviors of other people and generally ignore their own information which might cost them too much to obtain or analyze. Following this idea, we propose to detect herd behavior by focusing on investors' decision-making time variation. We observed that friend bids and bid counts impose significant effects on the decision-making time of investors, which is considered as the evidence of herding. We also conduct empirical analyses to address the impact of herd behavior on an individual's benefit. We reveal that lenders are more likely to herd on listings with more bids and friend bid, but their benefit will be reduced as the consequence of the behavior.

Keywords Herd behavior · Social network · Decision tree · P2P lending market

1 Introduction

With the innovative applicationa of IT technology in the financial industry, a new kind of lending market was inaugurated in 2005. The online peer-to-peer (P2P)

B. Luo (✉)
Sichuan Key Lab of Financial Intelligence and Financial Engineering,
Southwestern University of Finance and Economics, Chengdu 610074, Sichuan, China
e-mail: luobinjie@gmail.com

B. Luo · Z. Lin
Center for Advanced Analytics and Business Intelligence, Texas Tech University,
Lubbock, TX 79409, USA

lending market matches people who need small loans but are unable to get them from traditional lending markets, which are hosted primarily by banks and other financial institutions with willing lenders. The online P2P market allows these parties to communicate peer-to-peer directly. Without banks as intermediaries, the P2P lending market can theoretically bring significant benefits to both borrowers and lenders. The rapid growth and promising future of this market have also attracted considerable attentions (Everett 2010; Freedman and Jin 2008; Lopez et al. 2009; Lin et al. 2009; Iyer et al. 2009; Puro et al. 2010). However, there is a dearth of research on human behavior in the online P2P lending market. The emergence of Prosper.com, America's largest peer-to-peer lending market provides us with a good opportunity to fill this gap.

Herd behavior as a significant mode of human behavior has been documented in many academic fields: psychology (e.g., Tard and Parsons 1903; Hoffer 1955), financial markets (e.g., Brunnermeier 2001; Hirshleifer and Teoh 2003) and online auctions and purchasing (e.g., Dholakia and Bagozzi 2001; Chen 2008). Herd behavior describes various social situations in which individuals are strongly influenced by the decisions of others (Asch 1956). On a daily basis, most people appear to primarily use public information to infer the beliefs of others. Based on this interpretation, individuals form their own beliefs and make decisions. This tendency leads to herd behavior, where everyone follows what other people are doing regardless of the content of their private information (Banerjee 1992). For animals, herd behavior is considered as a good way of reducing the chance of being caught by a predator (Hamilton 1971). However, in human society, many cases have indicated that herd behavior is often correlated with undesired and inefficient results: herd behavior among investors has been believed to be one of the causes of the dot-com bubble in 1990s (Phillips et al. 2009); the "Tequila effect" refers to the herd behavior by global investors that resulted in a series of currency crises in South America in 1994 (Calvo and Mendoza 2000); and in online auctions, herd behavior may lead to neither the seller nor the buyer receiving the best value (Dholakia and Bagozzi 2001). Will herd behavior exist in the online P2P lending market? How could herd behavior form in such an environment? What is the impact of herd behavior on the investors exhibiting it? With data provided by the Prosper.com, this paper is intended to investigate these questions.

This paper is organized as follows. In Sect. 2, we conduct a review of the literature on herd behavior and P2P lending market. The details of our proposed model are presented in Sect. 3. We perform the empirical tests on the derived hypotheses in Sect. 4. In Sec. 5, we provide a summary and discussion about our work.

2 Background and literature review

In this section, we first introduce the process how a loan is obtained in Prosper.com. Based on the context, we review the prior research work on the online P2P lending market and those on herd behavior.

2.1 The loan formation on Prosper.com

The pricing model utilized by Prosper is a multi-unit reverse auction. Each potential borrower who wants to obtain money on Prosper has to register first. During the registration process, each potential borrower will be asked to provide his/her personal information as a means of identification. Then, Prosper will check and retrieve the person's credit score from a credit bureau (Experian). Based on this credit score, a Prosper Rating, which has seven levels from AA to HR in descending order, will be assigned to the person by directly translating his or her Experian credit score (e.g., a 700 credit score translates to a B credit grade). After registering, the borrower can create a customized listing, which is similar to a poster in a BBS (Bulletin Board System). This listing may include the borrower's personal and financial situation (e.g., sex, income, and job), a title and a short description regarding the purpose of the loan and the details of the borrower's request (e.g., the amount of money needed and the maximum interest rate that he or she is willing to accept). After creating a listing, the borrower can post it on the Prosper market and let lenders review it. At this point, the borrower has finished his or her work for obtaining a loan, and the market determines whether the listing can turn into a loan. Each listing is live for 14 days. During this period, the listing can be viewed by lenders (investor, lender, bidder, and decision maker are identical terms and interchangeable throughout the rest of this paper). After reviewing the listing, each lender will decide whether to bid. A listing can receive bids from different lenders, and the minimum amount of each bid is 50 dollars. Each lender who bids on a listing must reveal his or her interest rate and the amount of money that he or she is willing to lend to the borrowers. While the listing is live, if a lender failed in his or her previous bids, then the lender can change his or her asking interest rate and the amount of money that he or she is willing to lend to the owner of the listing. After doing so, the lender can bid again. Thus, based on how fierce the competition is, the interest rate may change dynamically during the bidding process and only settle down at closing time. If a listing is fully funded, then it will become a loan, and the funds will be deposited directly into the borrower's bank account within a few days. The borrower will also begin to make monthly payments to lenders per the agreement. Otherwise, the borrower can create another new listing and post it on Prosper again.

2.2 Research on P2P lending markets

It is widely accepted that financial intermediation exists primarily to mitigate the information asymmetry between borrowers and investors (Leland and Pyle 1977). As it is a new way of investing and borrowing money, the anonymity of the online P2P lending market may cause borrowers to exhibit greater uncertainty, thus worsening the problem of information asymmetry. Indirectly, it reduces investor's incentive to participate or may even lead to market failure, as described by Akerlof (1970). Social network services were introduced to the online P2P lending market to reduce the borrowers' level of uncertainty and to mitigate the problem of information asymmetry (Freedman and Jin 2008). Thus, one of the research directions in the P2P lending market is to investigate the effect of social network on the loan default rate. Everett

(2010) found that the loan default rate is significantly reduced only if membership holds the possibility of real-life personal connections. Similar evidence was reported by Freedman and Jin (2008). They said that loans with friend endorsements and friend bids have fewer missed payments and estimated returns of group loans are significantly lower than those of non-group loans. Additionally, Iyer et al. (2009) revealed that, on Prosper.com, the credit score is a reliable proxy for creditworthiness and should play an important role in the decision making of lenders. After analyzing the determinants affecting the success rate of obtaining a loan, Puro et al. (2010) developed a decision aid system for improving borrowers' decision quality on Prosper.com. To date, the research has examined the determinants in decision making for both lenders and borrowers in the online P2P lending market. However, the existence of Prosper.com also provides us with a chance to observe the trading behavior of traders in a new lending model that may complement the traditional lending model used by banks and other financial institutions. In this paper, we will investigate herd behavior on Prosper.com, which has not yet been addressed.

2.3 Models for herd behavior

In this section, we mainly review the economic literature that we consider relevant to our study. For a more extensive review of herd behavior not confined to the economic domain, we suggest reading Raafat et al. (2009). According to past theories, there are many reasons for us to behave like a herd, which have been used as a basis to develop different models. One of those reasons is "conformity preference": individuals inherently wish to conform to the behavior of others (Jones 1984). Based on this, Bikhchandani et al. (1992) proposed a model named "Information Cascades" to explain fads, fashions, customs, and cultural changes. Information cascades occur when the existing aggregated information becomes so overwhelming that an individual's private information is not strong enough to reverse the resulting decision. Almost simultaneously, Banerjee (1992) and Welch (1992) both modeled herd behavior as cascades. Their initial contributions inspired many similar models (Lee 1993; Banerjee and Fudenberg 1995; Brandenburger and Polak 1996) and formed one of the most important types of herd model.

Banerjee (1992) proposed an easy-to-understand model that we want to introduce here to illustrate how herd behavior happens. The model utilizes a very common situation in which people have to choose between two restaurants that are generally unknown. The preceding probability that the people will choose restaurant A is higher than that for B at 51%. People arrive at the restaurants in sequence, observe the choices made by the people before them, and decide on one of the two restaurants. In addition, each person also sends his or her own signal indicating a preference for A or B, and the quality of the signals is the same. Suppose that among 100 people, 99 receive signals indicating that B is better, but the one person whose signal favors A moves first. By observing the first person's action, the second person infers that his or her own signal favors A, even though the second person's signal actually favors B. Because these signals are equal in quality, they cancel out, and the rational choice for the second person is to follow the preceding probability and go to restaurant A. The second person exhibits herding behavior and selects A regardless

of his or her own signal. The third person faces the exact same situation as the second person. Because the second person's choice does not provide any new information to the next person, the third person then makes the same choice as the second person. Although B is better in general, everyone ends up going to A. In this simple model, the second person's decision, which ignores his or her own information and joins the herd, imposes a negative externality on the rest of the population. The latest model in this "cascade" stream was developed by Smith and Sorensen (2000). It models how Bayesian-rational individuals learn sequentially from the discrete actions of others, and it allows for individuals to possibly entertain different preferences over their actions, differentiating it from previous studies.

Another reason for herd behavior is reputation concern. The representative model was introduced by Scharfstein and Stein (1990). Based on the "The General Theory of Employment" (Keynes 1936), it explained why managers mimic others in investment decisions. The models of Trueman (1994), Graham (1999), Zwiebel (1995), and Prendergast and Stole (1996) also fall into this category but are devoted to different investment activities: analysts releasing forecasts (Trueman 1994; Graham 1999), managers adopting innovations (Zwiebel1995) and investment decisions (Prendergast and Stole 1996). Due to limited space and a lack of representativeness, we will not delve into the details of each of these models here.

Additionally, Froot et al. (1992) modeled how the trading horizon induces herd behavior. Avery and Chevalier (1999) presented a model predicting that older agents will herd less. Hirshleifer et al. (1994) demonstrated that the sequential nature of information arrival has a significant effect on herd behavior. Kultti and Miettinen (2007) modeled herd behavior with the cost of observation.

After reviewing those models, we find that they fail to fit the situations in the online P2P lending market because of the followings: (1) Most of the models assume that decision makers can obtain and study information for free except Kultti and Miettinen (2007). In our consideration, such kind of cost should play a key role in generating herd behavior in certain environments. (2) Most of previous models ignore individual information preference, although it may have an important effect on herding inclination (Welch 1992; Smith and Sorensen 2000). (3) They often assume that the succeeding decision makers observe perfect information from their predecessors. This is not applicable in the online P2P lending market, as only the predecessors who bid on a listing can be observed, while those who do not are not observed. This biased information display may impair the judgments of some inexperienced lenders. Being different from those reviewed models, we model the effect of personal preference on information and cost of obtaining and studying information on herd behavior simultaneously. Under the framework of decision tree, we analyze the optimal choice of decision maker and derive the conditions of herding. Based on that, we can see how information preference and costly information trigger herd behavior in an online P2P lending market.

2.4 Empirical works on herd behavior

As pointed out by Graham (1999), the best way to test herd behavior is to compare an individual's action with his private information because this would let us know

whether an individual discard his private information to take the public action. However, there seems to be no such data source that could match this demand. Instead, other ways have been developed to empirically test herd behavior. The most frequently used phenomenon to identify herd behavior is "action conformity": many people take the same action perhaps because some mimic the actions of others (Graham 1999). For example, by a observing a disproportionate share of investors who engage in buying, or at other times selling, the same stock, Lakonishok et al. (1992) concluded that the money managers in their sample did not exhibit significant herding behavior. Similar approaches were also employed by Peles (1993), Falkenstein (1996), Nofsinger and Sias (1996), and Wylie (1996) to test herd behavior among pension funds, mutual funds, and institutional investors.

The reasons behind this method are as follows: (1) An individual's private information or belief is hard to observe, but his/her actions are easy to capture; and (2) The occurrence of herd behavior must lead to "action conformity," which was suggested by many models (e.g., Bikhchandani et al. 1992; Banerjee 1992; Smith and Sorensen 2000; Welch 1992). However, the observed "action conformity" may not result from herd behavior. For instance, many investors purchase 'hot' stocks only due to correlated information arrival from independently acting investors (Devenow and Welch 1996). Grinblatt et al. (1995) and Wermers (1999) concluded that a large portion of herding behavior occurs when analysts "momentum follow," that is, when they buy recent winners or sell recent losers. If both a leader and a follower choose to momentum follow, then it can look like the latter is herding on the former, when in fact both are simply mimicking the market movement. Thus, the way to test herd behavior by observing "action conformity" is a compromise between theory and data availability.

Another deficiency of the previous empirical studies has been pointed out by Bikhchandani and Sharma (2001): empirical studies do not appear to be highly relevant to theoretical discussions; rather, the approach used in empirical studies is purely statistical. Lakonishok et al. (1992) utilized a purely statistical technique to measure investors' herd behavior that has since been applied in many papers (Grinblatt et al. 1995; Wermers 1999). There seems to be no work that makes a good connection between the theoretical and empirical discussions, except that of Graham (1999). Following the style of Graham (1999), we will first model the formation of herd behavior and then test the implications with data from Prosper.com.

3 Decision tree model

In this section, we use a decision tree to model the decision making process of investors and to explore how they herd on public information from Prosper.com. We depart from the previous models in the following ways: (1) There is a cost of information obtaining and studying. Profiles contain too much information, thus requiring a lender to spend time to read it and an additional click to access the details. However, bids by friends and the bid count are easier to access and understand for lenders. This is why we consider obtaining and studying profile to be a cost for lenders. (2) Integrating the concept of personal preference. In our model,

we allow different people to respond differently to the same information. (3) Closely connected to empirical tests. Unlike most of the theoretical models, the implications of our model are closely connected with hypotheses that will be verified by real data from Prosper.com.

3.1 Model settings

On a daily basis, people tend to infer others' signals or beliefs based on their observed actions (Frith and Frith 2008). The same thing happens on Prosper.com, where each lender can infer the quality of the borrower and update their beliefs by utilizing the available information, such as bids from predecessors or the borrower's friends. We assume that there is a sequence of lenders that will bid throughout the entire time period during which the listing is biddable. Each lender has to decide whether to bid. Simultaneously, the lender has to consider the interest rate that he/she is willing to accept. Nonetheless, the analysis of interest rates is not the focus of this study. For simplicity, we assume that each lender sets the interest rate as the same as that of his predecessors, which is an optimal and applicable strategy on Prosper.com.

We assume all of the lenders are rational, risk-neutral and expected payoff maximizers. On Prosper.com, we assume there are two types of borrowers: high (H) and low (L). If a borrower's type is H, then the payoff is V_H. Otherwise, the loss will be V_L. If a lender does not bid, then his or her payoff is 0. In Prosper, lenders have information to make an inference about the borrower type. There are at least three kinds of information that can be utilized by lenders: friend bid, bid count and borrower's profile. Lenders will study these types of information sequentially and update their beliefs about the borrower type in a step-by-step manner.

Assumption 1 Lenders are allowed to infer different beliefs with the same information.

When lender i and j observe the same information, such as a friend bid, as a consequence of assumption 1, we can say that lender i's belief about the borrower type is $P(Borrower = H|friend\,bid) = P_i$ which is not necessarily equal to lender j's belief of $P(Borrower = H|friend\,bid) = P_j$. We ascribe this to personal preferences for certain information. In contrast, in an information cascade model, Banerjee (1992) assumes that "signals are equal in quality," which means that people respond identically to the same information.

Assumption 2 The cost for obtaining and studying the information about the borrower's friend bid and bid count are minimal and negligible, while the cost for obtaining and studying the borrower's profile is significant.

This assumption arises from the fact that, on Prosper.com, information about a friend bid and bid count is easier to access and to understand. When a borrower's profile contains more than 20 items, including whether a borrower has a car or not, the past behavior of this borrower (e.g., whether this borrower owns a house), time and effort are needed to read it. Domain knowledge and experience are also required

to have a better understanding of profiles. Figure 2 is a snapshot of what a borrower's profile contains.

Because we have assumed that all lenders are rational, they will first study the easier information (friend bid and bid count) and then consider whether to obtain and study harder information (borrower's profile). The cost for obtaining and studying a borrower's profile is C.

Assumption 3 Lender's belief regarding a borrower's type will be affected by the number of bids on a listing. As the number of bids increases, the lender's belief that the borrower is of a high type will become stronger.

Based on assumption 3, we have $P(borrower = H|bid\ count = n)$ will increase as n increases but with a decreasing marginal effect until approximately 1.

3.2 Decision tree model

We model the decision process of lender i on Prosper.com, which is shown in Fig. 1.

Step 1 At the start point of step 1, lender i has a prior belief regarding the whole market before seeing any listings. The prior belief (p_0) may be given by nature or inferred from a public action like a market administrator's announcement.

$$\begin{cases} P(Borrower = H) = P_0 \\ P(Borrower = L) = 1 - P_0 \end{cases} \tag{1}$$

In step 1, lender i will probably observe a friend bid. If lender i sees it, he must update his belief from P_0 to P_1. If not, his belief will remain as P_0.

$$\begin{cases} P(Borrower = H|friend\ bid = yes) = P_1 \\ P(Borrower = L|friend\ bid = yes) = 1 - P_1 \end{cases} \tag{2}$$

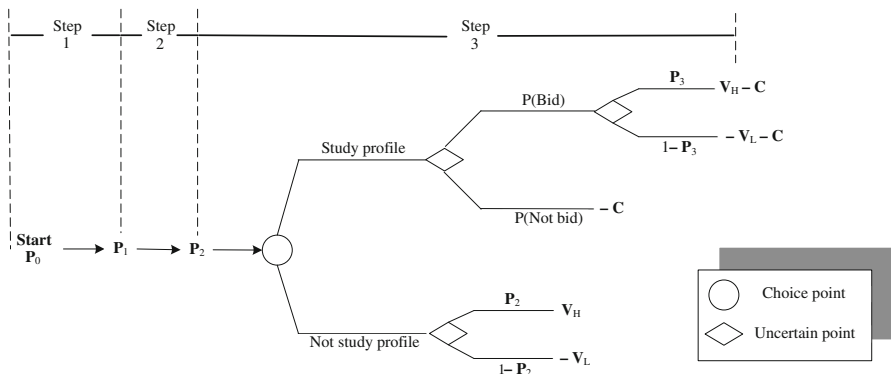*Step 2* In step 2, lender i will observe the others' bids, and his belief will be updated from P_1 to P_2.

Fig. 1 Decision tree for lender i

$$\begin{cases} P(Borrower = H | friend\,bid = yes,\, bid\,count = n) = P_2 \\ P(Borrower = L | friend\,bid = yes,\, bid\,count = n) = 1 - P_2 \end{cases} \quad (3)$$

As consequence of assumption 3, n increases, P_2 will increase. Additionally, if lender i favors more on the information of bid count, the positive effect of bid count might nullify any negative effect coming from previous information and P_2 can approximate 1. The current expected payoff is $EP_2 = P_2 \times V_H - (1 - P_2) \times V_L$.

Step 3 In step 3, lender i has to make a choice of whether to study a borrower's profile or not. If he chooses to do so, he has to pay a cost of C and the updated belief is:

$$\begin{cases} P(Borrower = H | friend\,bid = yes,\, bid\,count = n,\, profile) = P_3 \\ P(Borrower = L | friend\,bid = yes,\, bid\,count = n,\, profile) = 1 - P_3 \end{cases} \quad (4)$$

The conclusion inferred from a borrower's profile can be good, bad or neutral. If lender i considers it as a good profile, we should have $P_3 > P_2$. If not, $P_3 \leq P_2$. When $P_3 \times V_H - (1 - P_3) \times V_L > 0$, the optimal choice for lender i is to bid. The expected payoff for studying a profile is $EP_3 = P_3 \times V_H - (1 - P_3) \times V_L - C$. If the following conditions are satisfied,

$$\begin{cases} EP_{study} < EP_{not\,study} \\ P_3 \times V_H - (1 - P_3) \times V_L > 0 \end{cases} \quad (5)$$

the optimal choice for lender i is not to obtain and study the borrower's profile but to follow the public in deciding whether to bid. When this happens, we say that lender i behaves like the herd. Based on Eq. 5, we have the following propositions.

Proposition 1 *Keeping the other factors invariant, if the cost for obtaining and studying information increases, the probability of lender i to herd will also increase.*

Proposition 2 *Keeping the other factors invariant, if lender i has a strong preference for public information, he/she is more likely to herd.*

The proof is as follows. First, we consider EP_3 as a consequence of EP_2 and a random variable ϵ.

$$EP_3 = EP_2 + \epsilon - C \quad (6)$$

According to the decision tree model, we have $EP_{study} = P(bid) \times EP_3 - (1 - P(bid)) \times C$, and $EP_{not\,study} = EP_2$. Thus, the probability of herding is

$$\begin{aligned} P(herd) &= P(bid) \times P(EP_{study} < EP_{not\,study}) \\ &= P(bid) \times P(P(bid)) \times EP_3 - (1 - P(bid)) \times C < EP_2 \end{aligned} \quad (7)$$

Because $P(bid) = P(EP_2 + \epsilon > 0)$ and $EP_3 = EP_2 + \epsilon - C$, if we keep EP_2 invariant, as C increases, $P(bid)$ will stay the same, but $P(EP_{study} < EP_{not\,study})$ will increase. The probability to herd will increase along with the cost of further study, and proposition 1 is proved.

Based on Eq. 7, if we control C and let $EP_2 > 0$, we have

$$P(bid) \times (EP_2 + \epsilon - C) - (1 - P(bid)) \times C < EP_2$$
$$\Rightarrow P(bid)EP_2 + P(bid)\epsilon - P(bid)C - C + P(bid) \times C < EP_2 \qquad (8)$$
$$\Rightarrow P(bid)EP_2 + P(bid)\epsilon - C < EP_2$$

Because $P(bid) \leq 1$, as $\frac{|EP_2|}{|\epsilon|}$ increases (it is more likely to happen if assumptions three and four hold, which means lender i has a very strong preference for public information or the public information overwhelms as the bid count increases), $P(EP_{study} < EP_{not\ study})$ increases. In the meantime, $P(bid)$ will keep constant or increase. Therefore, we have proved proposition 2, in which lender i's preference for public information will lead to herd with a higher probability.

In real life, there are many causes that may lead people to have strong preference for public information, such as people not being confident with their information processing ability and being more willing to believe the information from experts or insiders. This kind of herding is similar to the reputation concern type herding (Scharfstein and Stein 1990): manager A has a stronger preference for manager B's information. Although A's evidence signals the opposite direction, he/she decides to herd.

Corollary 1 *The propositions 1 and 2 imply that, borrower's social relationship providing comparatively high-precision public information (friend bid) as well as the cost for obtaining and studying profile make lender's belief towards adopting public action. Therefore, lenders are more likely to herd on the free public information.*

Corollary 2 *Proposition 2 also implies that, an increased number of bids for a listing increases the probability of lenders' herding.*

Another question is how imperfect information affects the number of occurrences of herd behavior. Based on assumption 3, we derive that $P(Borrower = H|bid\ count = n)$ increases as n increases until approximately 1. If the perfect history about a listing is obtainable (the number of bids and the number of people who viewed a listing but without any bids), the speed of $P(Borrower = H|bid\ count = n)$ approaching 1 would not be as rapid as in the situation in which only the bid count is available. In other words, perfect information about a listing history would help to reduce the number of occurrences of lender herding on Prosper.com.

So far, we have modeled the formation of herding when lenders have heterogeneous preferences on information as well as costly information obtaining and studying. For the generalization of this model, we conjecture that it is not only confined to lenders in online lending market like Prosper, but is also applicable in predicating human behavior under other environments if assumption 1 and 2 are satisfied. At here, we give an illustrative example to see how to apply this model to other situation. Assume person A wants to buy a \$10 product at Walmart. When he was walking around in Walmart, he met one of his friends who said that he seemed to have seen the same product in another store is only \$8. Now, A is facing the cost

to obtain exact price information in another store and he should also have preference on his friend's words, such as it is reliable in 90% cases. If A has very strong belief that his friend is absolutely right, the only thing that would affect whether he will go to that store is the difference between travel cost and money saved by doing that. But in most of the cases, A's decision is determined by the benefit he will get by searching, which is based on his preference on his friend's words and search cost.

3.3 Hypotheses development

Based on corollaries 1 and 2, we propose several hypotheses for empirical tests. If free high-precision public information is available, lenders are less likely to pay the cost for obtaining and studying additional information. Their optimal choice is to herd on public information. If this happens, we have hypothesis 1.

Hypothesis 1 Due to herding effect, the average time interval between two consecutive bids is smaller for listings that contain a friend bid.

By considering that a friend bid can arrive at any time during which the listing is live, we propose another hypothesis.

Hypothesis 2 For listings with a friend bid, the average time interval between two consecutive bids is smaller after a friend bid appears.

By considering other potential determinants of average time intervals, we formulate hypotheses 3 and 4.

Hypothesis 3 After controlling for other potential determinants, the friend bid has a statistically significant effect on the average time interval between bids.

Hypothesis 4 After controlling for other potential determinants, the bid count has a statistically significant effect on the average time interval between bids.

In this section, we modeled how heterogeneous preferences and costly information affect the probability of herding in Prosper and in the next section, we will use the data downloaded from Prosper.com to test the proposed hypotheses.

4 Empirical results

4.1 Data description

We downloaded the data used to verify the proposed hypotheses in Sect. 3 from Prosper.com on 02/02/2010. The dataset contains listings and all the bids for each listing. The creation date of listings ranges from 11/09/2005 to 11/01/2010. For each listing, we have all the information displayed in Fig. 2. For each bid, we have the information on its creation date and time, bid amount, interest rate and the relationship between bidder and listing owner. The original dataset contained 925,130 listings and 6,550,387 bids. However, parts of the listings are heavily contaminated by noisy data and missing values. We removed listings with missing

Fig. 2 Items in a borrower's profile (see for the details http://www.prosper.com/help/topics/lender-listing-data.aspx)

and abnormal values. After that, the number of listings is 97,039 and the total number of bids for the 97,039 listings is 4,811,483. The average time interval for two consecutive bids was 7,873 s (i.e., approximately 2 h).

4.2 Verification of hypotheses

In hypothesis 1, we postulated that due to the herding effect, the average time interval between two consecutive bids is smaller for listings with a friend bid. To test this hypothesis, we divided the listings into two groups: a group composed of listings containing at least one bid from listing owner's friend, and a group composed of listings with no bids from friends. The statistical comparisons are presented in Table 1.

In Table 1, the average time interval for listings with a friend bid was 6,251 s (i.e., approximately 1.7 h), and the average time interval for listings without a friend bid was 2.2 h. Thus, the average time interval was 26.9% higher for the latter than for the former. The statistical significance of this difference also needed to be tested.

Table 1 Statistical comparison of listings with and without a friend bid

With friend bid	# of Listings	Average time interval	Range of time interval
Yes	1,952	6,251	[0–1,014,478]
No	95,087	7,931	[0–1,204,176]

We first checked whether the assumption of equal variance holds. The F-statistic of Levene's test for homogeneity of variance was 74.49, and the corresponding p-value was less than 0.05. Thus, homogeneity variance is rejected. Under heterogeneous variance, the Welch's t test is reported: the t-value is 21.16, and the p-value is less than 0.05. The difference in the average time interval between the two groups is statistically significant.

Result 1   *The average time interval for listings with a friend bid is shorter than that for listings without a friend bid at a statistically significant level. The introduction of a friend bid can shorten the time interval between bids, suggesting that lenders are more likely to rely on high-precision public information than to spend time screening listings by themselves. Thus, lenders behave like a herd.*

We proposed hypothesis 2 because a friend bid can arrive at any time while a listing is live. With respect to the time stamp of the first friend bid, we divided the bids for each listing with a friend bid into two groups: a group composed of bids that were placed before the first friend bid and a group composed of bids placed after the first friend bid. A statistical comparison is listed in Table 2.

Following the same procedure used to verify hypothesis 1, we check whether the assumption of homogeneity variance holds first. The F statistic of Levene's test is 4.73, and the corresponding p-value is 0.0296. The results suggest we reject it. Thus, we use Welch's test to make sure whether the difference is statistically significant. The t-value of Welch's t test 2.43 and its p-value is 0.0151. From Table 2, we can see that the introduction of a friend bid has reduced the time interval between two consecutive bids by 12.5%. After a friend bid appeared, the average time interval between two consecutive bids was smaller. Because this difference is statistically significant, hypothesis 2 has been proved.

Result 2   *Friend bid plays a key role in triggering a lender's herd behavior.*

To verify hypotheses 3 and 4, we choose part of factors (listed in Table 3) from Fig. 2 according to their relevance to our research questions as well as those ones that have been proved to have significant effects on listings' success rate by previous studies (e.g., Puro et al. 2010; Freedman and Jin 2008; Lin et al. 2009).

Table 2 Statistical comparison of average time interval before and after friend bid

After the first friend bid	# of Bids	Average time interval	Range of time interval
Yes	127,066	5,753.4	[0–663,587]
No	38,766	6,574.5	[0–790,211]

Table 3 Variables and descriptions

Variable	Description
BidCount	The number of bids that a listing received during its available time period
FriendBid	A binary variable indicating whether a bid was placed by a friend of the listing owner (i.e., borrower). If so, then FriendBid = 1. Otherwise, Friendbid = 0
RequestedAmount	Listing requested amount
MaximumInterestRate	Maximum interest rate set by the borrower
CreditGrade	Credit grade of the borrower including 7 levels from AA to HR in descending order and NC indicates not being assigned to any credit level
DebtToIncomeRatio	Debit to income ratio of listing owner
FundingOptions	The funding options is one of the following values: open for duration—the listing is open for it's duration. Close when funded—the listing will close as soon as it is funded 100%
HasGroup	Whether or not a listing owner joined a group. If joined, HasGroup = 1. Otherwise, HasGroup = 0
HasImage	Whether or not a listing contains an image. If has, HasImage = 1 otherwise 0
HasHome	Whether or not the listing owner has is a verified Homeowner at the time the listing was created. If has, HasHome = 1 otherwise 0
ProsperRating	Prosper rating at the time the listing was created. It includes 7 levels from AA to HR in descending order and NC indicates not being assigned to any level
Category	The category of this listing. The category is one of the following numerical values:
	0 Not available
	1 Debt consolidation
	2 Home improvement loan
	3 Business loan
	4 Personal loan
	5 Student loan
	6 Auto loan
	7 Other
Endorsements	Whether or not the listing owner get endorsements. If has, endorsements = 1 otherwise 0

We use a regression model (Eq. 9) controlling other potential determinants to verify hypotheses 3 and 4. All of the coefficients of regression models in this paper are estimated by the GLS (generalized least squares) method, which can produce unbiased estimates in the presence of conditionally heteroskedastic and serially correlated errors.

$$\text{Average time interval}_i = \beta_0 + \sum \beta_j x_i + \epsilon_i \quad \text{where } \epsilon_i \in N(\mu, \sigma^2) \qquad (9)$$

The R-square of Eq. 9 is 21.09%, and the adjusted R-square is 20.46%. The estimated coefficients are presented in Table 4.

Result 3 *After controlling for the other important factors, friend bid and bid count still exhibit significant effects on the average time interval. Friend bid and bid count*

can significantly reduce the length of the average time interval which indicate that lenders are more likely to herd on listings with friend bid and more bids.

To compare the weights of the different effects on the average time interval, we list the Type 3 sums of squares for each effect in Table 5.

Result 4 *Bid count and friend bid are the top 2 important factors affecting average time interval. They contribute 91% to the explained variance. According to Table 7, the number of bids is the most important variable in determining the average time interval and friend bid is less important than it.*

4.3 The impact of herd behavior on the lender's benefit

As suggested by Simonsohn and Ariely (2008) and Dholakia and Soltysinski (2001), bidders herd on auctions with more existing bids and, consequently, ignore comparable or even better choices. Dholakia and Soltysinski (2001) named this phenomenon "herding bias". Here, we argue that because lenders tend to herd on listings with more bids, inefficient outcomes may occur. To prove this argument, we first define two terms: the potential loss benefit (PLB), which is the difference between the maximum interest rate set by the borrower and the last traded interest rate, and the intensity of competition, which is the ratio of the number of bids to the requested amount. According to our findings, increasing the number of bids increases the probability of herding behavior. Thus, lists with a higher intensity of competition will also exhibit a higher probability of herding. We estimated Eq. 10 for the listings that had been labeled as "Paid" by Prosper.com. "Paid" listings indicate that the loans have been fully paid. The R-square was 46.64%, the adjusted R-square was 46.59%, and the estimated coefficients are listed in Table 6. The predicted regression lines for the top 3 credit grades are also shown in Fig. 3. Because the other five lines are too close to each other and cannot be distinguished from one another in a single figure, they have been ignored.

$$\text{PLB}_i = \beta_0 + \beta_1 CG_i + \beta_2 Ratio_i + \epsilon_i \quad \text{where } \epsilon_i \in N\left(\mu, \sigma^2\right) \qquad (10)$$

Based on Table 6; Fig. 3, we find that an increase in the ratio, which represents the intensity of competition for a listing, will result in a greater loss of profit while retaining the "Paid" status of the loan.

Result 5 *Herding on listings with more bids might reduce lenders' benefits.*

In this section, we proposed a new way to detect herd behavior by focusing on the change in the time interval between two consecutive bids made by lenders on Prosper.com. We observed that a friend bid and the bid count impose significant effects on the decision-making time of lenders, which is considered as evidence of herding. We also revealed that herding on listings with more bids impairs the lender's benefit. We have two considerations about the generalization of this

Table 4 Parameter estimates for Eq. 9

Parameter	DF	Estimate	Standardized estimate	Standard error	t value	Pr > |t|
Intercept	1	10.3893	0	0.7261	14.31	<0.0001
BidCount	1	−0.0074	−0.4379	0.0003	−25.78	<0.0001
FriendBid = 1	1	−0.5022	−0.1194	0.0806	−6.23	<0.0001
RequestedAmount	1	0.0444	0.0176	0.0426	1.04	0.2973
MaximumInterestRate	1	1.6323	0.0648	0.4817	3.39	0.0007
CreditGrade = AA	1	−0.8128	−0.1050	0.6471	−1.26	0.2091
CreditGrade = A	1	−0.8106	−0.1126	0.6435	−1.26	0.2079
CreditGrade = B	1	−0.7697	−0.1173	0.6398	−1.20	0.2290
CreditGrade = C	1	−0.7249	−0.1318	0.6357	−1.14	0.2542
CreditGrade = D	1	−0.6216	−0.1169	0.6339	−0.98	0.3269
CreditGrade = E	1	−0.5614	−0.0861	0.6351	−0.88	0.3768
CreditGrade = HR	1	−0.6397	−0.1253	0.6318	−1.01	0.3114
DebtToIncomeRatio	1	−0.0200	−0.0121	0.0239	−0.84	0.4018
FundingOptions = CloseWhenFunded	1	−0.0138	−0.0025	0.0846	−0.16	0.8706
HasGroup = 1	1	−0.2246	−0.0504	0.0707	−3.17	0.0015
HasImage = 1	1	−0.2367	−0.0543	0.0652	−3.63	0.0003
HasHome = 1	1	−0.1654	−0.039049	0.065478	−2.53	0.0116
ProsperRating = AA	1	0.81354	0.0314	0.3888	2.09	0.0365
ProsperRating = A	1	0.1959	0.0104	0.2797	0.70	0.4837
ProsperRating = B	1	−0.1315	−0.0036	0.5286	−0.25	0.8035
ProsperRating = C	1	0.6698	0.0278	0.3515	1.91	0.0567
ProsperRating = D	1	−0.0325	−0.0013	0.3564	−0.09	0.9274
ProsperRating = E	1	0.0912	0.0029	0.4622	0.20	0.8437
ProsperRating = HR	1	0.6111	0.0348	0.2614	2.34	0.0194
Category = 0	1	−0.0564	−0.0134	0.1471	−0.38	0.7010
Category = 1	1	0.0932	0.0195	0.1421	0.66	0.5119
Category = 2	1	−0.1466	−0.0108	0.2316	−0.63	0.5267
Category = 3	1	0.1271	0.0164	0.1687	0.75	0.4514
Category = 4	1	−0.1412	−0.0184	0.1670	−0.85	0.3979
Category = 5	1	0.1407	0.0092	0.2529	0.56	0.5782
Category = 6	1	−0.1404	−0.0104	0.2302	−0.61	0.5418
Endorsements = 1	1	−0.0441	−0.0105	0.0810	−0.55	0.5855

approach to identify herd behavior in other situations if decision makers are facing costly information and have heterogeneous preferences on information. First, an experimental study on potential factors for human decision-making time variation is applicable and needed. Second, this approach can be easily applied on other real data if the decision-making time is available.

Table 5 Type 3 sums of squares

Source	DF	Type 3 SS	Mean square	F value	Pr > F
BidCount	1	2,338.8977	2,338.8977	664.52	<0.0001
FriendBid	1	136.6012	136.6012	38.81	<0.0001
ProsperRating	7	46.6291	6.6613	1.89	0.0666
HasImage	1	46.3534	46.3534	13.17	0.0003
MaximumInterestRate	1	40.4137	40.4137	11.48	0.0007
HasGroup	1	35.4689	35.4689	10.08	0.0015
Category	7	27.3421	3.9060	1.11	0.3537
HasHome	1	22.4466	22.4466	6.38	0.0116
CreditGrade	7	17.3853	2.4836	0.71	0.6673
RequestedAmount	1	3.8236	3.8236	1.09	0.2973
DebtToIncomeRatio	1	2.4747	2.4747	0.70	0.4018
Endorsements	1	1.0470	1.0470	0.30	0.5855
FundingOptions	1	0.0935	0.0935	0.03	0.8706

Table 6 Parameter estimates for Eq. 10

Parameter	DF	Estimate	Standardized estimate	Standard error	t Value	Pr > |t|
Intercept	1	−0.0051	0	0.0053	−0.95	0.3400
Ratio	1	1.3709	0.6900	0.0115	119.27	<.0001
CG = AA	1	−0.0120	−0.1447	0.0053	−2.26	0.0236
CG = A	1	−0.0085	−0.0906	0.0053	−1.59	0.1110
CG = B	1	−0.0028	−0.0303	0.0053	−0.52	0.6024
CG = C	1	0.0016	0.0193	0.0053	0.30	0.7609
CG = D	1	0.0032	0.0368	0.0053	0.61	0.5437
CG = E	1	0.0024	0.0226	0.0053	0.46	0.6491
CG = HR	1	0.0029	0.02370	0.0053	0.54	0.5884
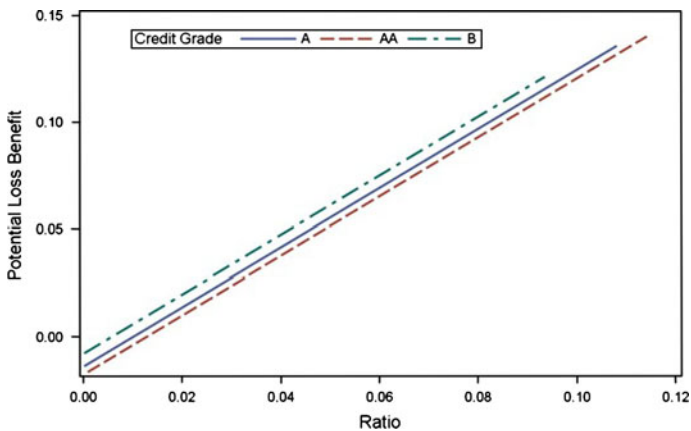
Fig. 3 Regression lines for Eq. 10

Table 7 Evidences for research questions

Research question	Hypothesis	Evidence
Will herd behavior exist in the online P2P lending market?	Hypotheses 1–4	Results 1–4
What is the impact of herd behavior on the investors exhibiting it?	Hypotheses 1–4	Result 5

5 Conclusions and discussions

Until now, we have provided explanation for the formation of herding in online P2P lending market with social network services, which can be considered as an answer to research question: How could herd behavior form in such an environment? Besides that, we also identified the existence of herding in Prosper and investigated the impact of herding on lender's benefits. Table 7 is a short summary about research questions and corresponding empirical evidences.

Scholars understand that online social networks represent a new form of social relationships, but little is known about how people behave in these networks. In this paper, we investigated herd behavior on Prosper.com, an online P2P lending market that incorporates social network services. With a decision tree, we modeled the decision making process of investors allowing heterogeneous preferences on information and costly information obtaining and studying. Unlike those found in previous theoretical works, we have derived several hypotheses from our model that are closely connected to the confirmation of herd behavior on Prosper.com. The social network service provided by Prosper.com gives each lender a chance to bid on his or her friend's listing. If lenders have very strong preference on public information as well as the cost for obtaining and studying additional information, they are more likely to behave like herd. Additionally, according to our empirical study, herding on listings with more bids reduces the lender's benefit. Because of that, we suggest that the market administrator should provide perfect information about a listing's history because only displaying the number of bids would cause inexperienced lenders to make an unbiased judgment. The availability of perfect information of the listing history would greatly reduce the probability of herding. For example, let there be two listings, A and B, both having received 100 bids already, but one has been viewed 100 times, and the other has been viewed 1000 times. For any lender, it is difficult to tell the difference in quality of those two listings from the bid count. However, lenders can easily locate the listing that is more favored by the public if the number of bids and the number of view times are both displayed. From an empirical standpoint, we considered the change in average time interval between two consecutive bids as the evidence of herding. We found that friend bid and the bid count affect time intervals in a statistically significant manner. This finding shows that, if possible, lenders are more likely to rely on public information to make their decisions, and they will occasionally exhibit herding behavior by refusing to study information that comes at a cost. In this study, we have conducted a comprehensive investigation of herding behavior within an online social network, Prosper.com. However, we think that the proposed model and the derived hypotheses regarding decision-making time variation in the P2P lending context still need further test in other contexts of study in order to be generalized.

References

Akerlof G (1970) The market for 'Lemons': quality uncertainty and the market mechanism. Quart J Econ 84(3):488–500

Asch SE (1956) Studies of independence and conformity: a majority of one against a unanimous majority. Psychol Monogr 70(9):70–79

Avery CA, Chevalier JA (1999) Herding over the career. Econ Lett 63:327–333

Banerjee AV (1992) A simple model of herd behavior. Q J Econ 107(3):797–817

Banerjee AV, Fudenberg D (1995) Word-of-mouth learning. Working paper, Department of Economics, Massachusetts Institute of Technology

Bikhchandani S, Sharma S (2001) Herd behavior in financial markets. IMF Staff Papers 47(3):279–310

Bikhchandani S, Hirshleifer O, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. J Political Econ 100:992–1026

Brandenburger A, Polak B (1996) When managers cover their posteriors: making the decisions the market wants to see. Rand J Econ 27:523–541

Brunnermeier MK (2001) Asset pricing under asymmetric information: bubbles, crashes, technical analysis, and herding. Oxford University Press, Oxford

Calvo GA, Mendoza EG (2000) Rational herd behavior and globalization of securities markets. J Int Econ 51(1):79–113

Chen Y (2008) Herd behavior in purchasing books online. Comput Hum Behav 24:1977–1992

Devenow A, Welch I (1996) Rational herding in financial economics. Eur Econ Rev 40:603–615

Dholakia UM, Bagozzi RP (2001) Consumer behavior in digital environments. In: Wind J, Mahajan V (eds) Digital marketing: global strategies from the world's leading experts. Wiley, New York, pp 163–200

Dholakia UM, Soltysinski K (2001) Coveted or overlooked? The psychology of bidding for comparable listings in digital auctions. Mark Lett 12(3):223–235

Everett CR (2010) Group membership, relationship banking and loan default risk: the case of online social lending. Available at SSRN: http://ssrn.com/abstract=1114428

Falkenstein E (1996) Preferences for stock characteristics as revealed by mutual fund portfolio holdings. J Finance 51:111–135

Freedman S, Jin GZ (2008) Do social networks solve information problems for peer-to-peer lending? Evidence from Prosper.Com. Available at SSRN: http://ssrn.com/abstract=1304138

Frith CD, Frith U (2008) Implicit and explicit processes in social cognition. Neuron 60:503–510

Froot KA, Scharfstein D, Stein JC (1992) Herd on the street Informational inefficiencies in a market with short-term speculation. J Finance XLVII 4:1461–1484

Graham JR (1999) Herding among investment newsletters: theory and evidence. J Finance LIV(1):237–268

Grinblatt M, Titman S, Wermers R (1995) Momentum investment strategies, portfolio performance, and herding: a study of mutual fund behavior. Am Econ Rev 85(5):1088–1105

Hamilton WD (1971) Geometry for the selfish herd. J Theor Biol 31:295–311

Hirshleifer D, Teoh SH (2003) Herd behavior and cascading in capital markets: a review and synthesis. Eur Financ Manag 9(1):25–66

Hirshleifer D, Subrahmanyam A, Titman S (1994) Security analysis and trading patterns when some investors receive private information before others. J Finance 49:1665–1698

Hoffer E (1955) The passionate state of mind. Harper and Row, New York

Iyer R, Khwaja AI, Luttmer EFP, Shue K (2009) Screening in new credit markets: can individual lenders infer borrower creditworthiness in peer-to-peer lending? Working paper, National Bureau of Economic Research

Jones SRG (1984) The economics of conformism. Blackwell, Oxford

Keynes JM (1936) The general theory of employment, lnterest and money. Macmillan, London

Kultti K, Miettinen P (2007) Herding with costly observation. BE J Theor Econ 7(1):1–14

Lakonishok J, Shleifer A, Vishny RW (1992) The impact of institutional trading on stock prices. J Financ Econ 32:23–43

Lee IH (1993) On the convergence of informational cascades. J Econ Theory 61:395–411

Leland HE, Pyle DH (1977) Informational asymmetries, financial structures, and financial intermediation. J Finance 32:371–387

Lin M, Prabhala NR, Viswanathan S (2009) Can social networks help mitigate information asymmetry in online markets? In: Proceedings of the 30th international conference on information systems, Phoenix, Arizona, USA

Lopez SH, Pao ASY, Bhattacharrya R (2009) The effects of social interactions on P2P lending. Working paper, MAS final project

Peles N (1993) The determinants of institutional demand for equity: an empirical study. Working paper, Columbia University

Nofsinger J, Sias R (1996) Herding by institutional and individual investors. Working paper, Marquette University

Phillips, PCB, Wu Y, Yu J (2009) Explosive behavior and the NASDAQ bubble in the 1990s: when did irrational exuberance escalate asset values? Int Econ Rev, in press. Also available at SSRN: http://ssrn.com/abstract=1413830

Prendergast C, Stole L (1996) Impetuous youngsters and jaded old-timers: acquiring a reputation for learning. J Political Econ 104:1105–1134

Puro L, Teich JE, Wallenius H, Wallenius J (2010) Borrower decision aid for people-to-people lending. Decis Support Syst 49:52–60

Raafat RM, Nick C, Frith C (2009) Herding in human. TRENDS Cogn Sci 13(10):420–428

Scharfstein DS, Stein JC (1990) Herd behavior and investment. Am Econ Rev 80(3):465–479

Simonsohn U, Ariely D (2008) When rational sellers face nonrational buyers: evidence from herding on eBay. Manag Sci 54(9):1624–1637

Smith L, Sorensen P (2000) Pathological outcomes of observational learning. Econometrica 68:371–398

Tard Gd, Parsons EWC (1903) The laws of imitation, H. Holt and Company, New York

Trueman B (1994) Analyst forecasts and herding behavior. Rev Financial Stud 7:97–124

Welch I (1992) Sequential sales, learning, and cascades. J Finance 47(2):695–732

Wermers R (1999) Mutual fund herding and the impact on stock prices. J Finance 54:581–622

Wylie S (1996) Tests of the accuracy of measures of herding using UK data. Working paper, London School of Business

Zwiebel J (1995) Corporate conservatism and relative compensation. J Political Econ 103:1–25