Dynamics of Bidding in a P2P Lending Service: Effects of Herding and Predicting Loan Success

Simla Ceyhan*
Stanford University
simlac@stanford.edu

Xiaolin Shi*
Stanford University
shixl@stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

Online peer-to-peer (P2P) lending services are a new type of social platform that enables individuals borrow and lend money directly from one to another. In this paper, we study the dynamics of bidding behavior in a P2P loan auction website, Prosper.com. We investigate the change of various attributes of loan requesting listings over time, such as the interest rate and the number of bids. We observe that there is herding behavior during bidding, and for most of the listings, the numbers of bids they receive reach spikes at very similar time points. We explain these phenomena by showing that there are economic and social factors that lenders take into account when deciding to bid on a listing.We also observe that the profits the lenders make are tied with their bidding preferences. Finally, we build a model based on the temporal progression of the bidding, that reliably predicts the success of a loan request listing, as well as whether a loan will be paid back or not.

Categories and Subject Descriptors: H.4 [Information Systems Applications]: Miscellaneous **General Terms:** Economics, Experimentation **Keywords:** Peer-to-peer lending service, user behavior, auction, dynamics

1. INTRODUCTION

Online peer to peer lending platforms such as Prosper [19], Kiva [14] and Lending Club [16] directly connect individuals who want to borrow money to those who want to lend it. These platforms eliminate the need for a financial institution as an intermediary between lenders and borrowers, with the consequence that the loss resulting from defaulting on a loan is directly borne by the lenders themselves. The risk is distributed among the lenders proportional to the amount of money they lend to a given borrower. The lending process typically involves an auction on the loan's interest rate: lenders who provide the lowest interest rates win the bids, i.e. they get to lend the money to the borrower. In this sense, the interest rate is an analogy to the price a bidder offers in the auction, and the amount of money a lender offers is an analogy to the risk a bidder is willing to take.

Communities such as Prosper create a social structure and interactions that are interesting from many perspectives. On the one hand, lenders have at their disposal a very rich set of data that could potentially affect their decision making. They have access to financial information concerning borrowers, such as credit scores, past borrowing histories, and demographic indicators (race, gender, and location). On the other hand there is a non-trivial risk of loan default, and as such the lenders face a clear trade-off between interest rates and the amounts they bid. Finally, P2P lending platforms fill a niche for borrowers who cannot get a loan from traditional financial institutions or who need small personal loans, and are expected to become increasingly popular given the current economic climate. A recent study [10] predicts that within the next three years, peer-to-peer lending will increase by 66% to a total volume of 5 billion USD in outstanding loans in 2013.

The loan auction mechanism for such a community works as follows: a borrower posts a loan with the amount of money she wants to borrow and the maximum interest rate she is willing to accept. Lenders submit their minimum interest rate as well as the amount of money that they wish to lend to the borrower. The listing lasts for a pre-determined duration of time. At the end of this time, the auction is either successful, if enough money has been bid on the listing, or is void if there is not enough money received by it. For the successful listings, the P2P lending platform then combines the bids with the lowest interest rates into a loan to the borrower and takes care of the collection of money and repayments.

In this paper, we study the dynamics of such bidding mechanism in the case of P2P lending provider Prosper Loans Marketplace. We investigate the factors that affect the bidding process throughout a listing's lifetime. We observe that bids for a single listing do not occur uniformly over the listing's lifetime. There is a clear concentration of bids at the beginning and at the end of a listing's life, as well as at the point where the total requested amount of money is about to be satisfied. We explain this phenomenon by showing that there are three main economic factors that lenders take into account when deciding to bid on a listing: lenders' belief about the probability of a listing being fully funded, the probability of winning the bid, as well as the interest rate. In addition, a social factor that makes lenders prefer for new listings also takes effect. These factors change over the lifetime of a listing, and result in the non-uniform distribution of bids over time.

We also examine the performance of individual lenders and observe that it is related to their bidding preferences. Lenders who bid at the end of the listings are more likely to win the bids. Lenders who bid around the time when the requested amount by the borrower is satisfied are least likely to win the bid and less likely to make profits. Moreover, the strategy of lenders minimizing their risks by decentralizing does only help mildly.

Finally, we build logistic regression models to predict the listing success. Given our previous observations regarding the non-uniform temporal bidding behavior, it is not surprising that the listing bid trajectory plays an important role on these models. Only based on the temporal progression of the bidding behavior, we

*Authors contributed equally to this work.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0632-4/11/03.

can well estimate if a listing gets funded or not as well as predict whether its borrower will pay it back or not. We show that there is information to be gained from “how the market feels" that is not present in the set of features constructed from standard factors such as credit grade or debt-to-income ratio. To the best of our knowledge, this is the first study that also focuses on the prediction of the performance of loans in addition to fundability success.

The rest of the paper is organized as follows: Section 2 summarizes the related research in peer-to-peer lending marketplaces. Section 3 briefly describes the dataset and basic statistics about Prosper. Section 4 explores the dynamics of bidding on a listing and looks at the performance of individual lenders while Section 5 describes a model that predicts a listing's success and being paid-back or not. Conclusions are discussed in Section 6.

2. RELATED WORK

Having emerged recently, dynamics of peer-to-peer lending has been relatively unexplored. So far, one line of research has focused on understanding the general dynamics of the online peer-to-peer lending marketplace. Hulme and Wright [12] provide a study of peer-to-peer lending focusing on “Zopa.com". Berger and Gliesner [3], with a more general approach, analyze the role of intermediaries on electronic marketplaces. Freedman and Jin [7], examine the functioning of online lending based on Prosper's transaction data. They study the effect of social network in identifying risks and find evidence both for and against it.

Prosper market data has been freely available via an API. This allowed researchers study the Prosper market and consumer credit markets in general. Using Prosper data, Freedman and Jin [8] look at the change of borrower and lender behavior over time as new policies were introduced. Iyer et al. [13] examine how well lenders on Prosper use information, both traditional and non-traditional, to infer a borrower's actual credit scores. They show that while lenders mostly rely on standard banking variables to draw inferences on creditworthiness, they also use non-standard, subjective sources of information in their screening process, especially in the lower credit categories. Similarly, Klafft [15] focuses on lender behavior and demonstrates that careful lenders who choose their borrowers in accordance with a number of easy-to-observe selection criteria can still expect their investments to be profitable. Our work also analyzes Prosper data in order to understand online peer-to-peer lending throughly. However, we focus on the dynamics of bidding behavior and investigate the change of several attributes of the loan request listing during the auction duration.

Another line of research looks at the determinants of success in online peer-to-peer communities. Analyzing Prosper data, Herzenstein et al. [11], found that borrowers' financial strength and efforts when they post a listing are major factors in determining whether it will be funded or not. Similarly, Ryan et al. [20], analyze fundability determinants in Prosper. The purpose of this study is to weight the relative relevance of each of the financial and social features independently, and determine their influence in the success on the conversion of a listing to a loan. Their study shows that financial features are determinant. In this paper, we also build a regression model to predict fundability success. However, rather than using borrower related features in our model, we also explore temporal dynamics of a listing and show that they reliably predict the success of a listing.

On the other hand, Chen et al. [4] analyze the mechanisms of social lending from a theoretical standpoint. Again focusing on Prosper, they show that its mechanism is exactly the same as the VCG mechanism applied to a modified instance of the problem. They provide a complete analysis and characterization of the Nash equilibria of the Prosper mechanism and show that while the Prosper mechanism is a simple uniform price mechanism, it can lead to much larger payments for the borrower than the VCG mechanism.

Our work is also related to empirical studies on Ebay, where researchers try to understand the bidding characteristics. For example, Shah et al. [21] focus on individual users and monitor all of a bidder's engagements in the dataset. They identify the main bidding strategies such as “late bidding" (bidding during the last hour), “evaluator" (bidding the true valuation early) and “sceptic" (bidding only the minimum required amount) using rule-based methods. Lucking-Reiley, et al. [18] analyze the effect of various eBay features on the final price of auctions. They find that seller's feedback rating have a measurable effect on his auction prices, with negative comments having a much greater effect than positive comments. In another study, Dietrich et al. [5] empirically show that the segmentation of the eBay marketplace affects bidding behavior. They state that successful strategies for a certain product category or seller type can be useless in different market segments. In our work, with a similar approach to Prosper data, we also want to understand the bidding dynamics and Prosper features that affect bidding behavior. However, Prosper mechanism, with multiple winners, is different from eBay auctions in which a single bid at or above the seller's reservation price results in a successful auction.

Finally, our work is similar in spirit to the empirical studies on bidding dynamics in sponsored search. Most of the academic literature on sponsored search is theoretical in nature, characterizing behavior or payoffs of bidders. However, there are some studies, similar to our work, that examine actual bidding data on a large scale to determine how real-world auctions can best be analyzed and understood. One example is the study of Asdemir [1], where he analyzes how advertisers bid for search phrases in pay-per-click search engine auctions. Similarly, by examining sponsored search auctions run by Overture (now part of Yahoo!) and Google, Edelman and Ostrovsky present evidence of strategic bidder behavior in these auctions in [6]. In [2] Auerbach et al. empirically investigate whether advertisers are maximizing their return on investment across multiple keywords in sponsored search auctions and in [9], they utilize sponsored search data drawn from a wide array of Overture/Yahoo! auctions and examine how bids are distributed and the evidence for strategic behavior.

3. PROSPER MARKETPLACE

After creating a personal profile, borrowers and lenders become members of the Prosper community. When a borrower wants to request a loan through the marketplace, she creates a listing for a specific amount and sets a maximum interest rate that she is willing to pay. She also chooses the duration for which the request will remain active. Each loan request, from now on we will refer to as ‘listing’, includes information about the borrower such as her current credit rate and debt-to-income ratio. This information is verified by Prosper and made public to potential lenders. Lenders bid for the privilege of supplying all or part of the requested loans specifying the amount and interest rate at which they are willing to lend. When the specified time has elapsed, if the aggregate amount offered by lenders exceeds the amount requested by the borrower, the listing becomes successful. For the successful listings, the bids with the lowest rates “win” and are combined into a single loan to the borrower, with Prosper acting as the broker between the lenders and the borrower.

Bidding on Prosper is done in a Dutch auction format, which means that multiple lenders can get a piece of the same loan in varying amounts. In a Dutch auction, the borrower starts the auction with a maximum interest rate and multiple lenders bid that rate

	Funded	Non-funded
Frac. of Amount Collected	1	0.054 (0.15)
Bid Count	135.1 (142.6)	6.4 (34.6)
Duration(days)	7.61 (2.0)	7.45 (2.0)
Borrower Rate	0.213 (0.07)	0.195 (0.08)
Lender Rate	0.183 (0.07)	0.192 (0.08)
Amount Requested	\$6,126 (5587)	\$7,541 (6383)
Debt to Income Ratio	0.16 (0.15)	0.23 (0.45)

Table 1: Data statistics of funded and unfunded listings: mean, and in parentheses, standard deviation.

	Paid	Not-paid
Bid Count	124.7 (137.8)	121.5 (144.0)
Duration(days)	7.5 (2.0)	7.67 (2.0)
Borrower Rate	0.178 (0.07)	0.242(0.06)
Lender Rate	0.154 (0.06)	0.216 (0.06)
Amount Requested	\$ 5,670 (5350)	\$6,573 (6171)
Debt to Income Ratio	0.28 (0.92)	0.38 (1.12)

Table 2: Data statistics of paid and not-paid listings: mean, and in parentheses, standard deviation.

down until the auction times out. Prosper provided us with data that contains all the bidding and membership data from November 2005 to August 2009. The data encloses approximately 5 million bids, 900,000 members and 350,000 listings. At the end of summer 2009, Prosper introduced automated plan system, which bids on behalf of the lenders once a listing that matches their plan is posted. However, note that our data is from the period before this feature was introduced and thus it only consists of bids by individual lenders.

On average, every borrower posts 1.7 listings and every lender bids on 2.6 unique listings. There are 24,295 successful (funded) listings that have ended up in loans, which is about 8% of all listings. Out of those listings, 70% of them had competition, by which we mean that they continued receiving bids even after the amount requested was satisfied. While 7668 (32%) of the loans were paid back, 7595 (31%) of them defaulted. For the rest, payoff is in progress. Table 1 shows the mean and, in parentheses, the standard deviation of a number of statistics related to both funded and non-funded listings. Funded listings have much higher number of bids, as a result, the average percentage of the amount collected by the non-funded listings is quite low, only 5%. Not surprisingly, on average funded listings have higher starting interest rates (21.3% vs. 19.5%), lower final interest rates (18.3% vs. 19.2%) and lower requested amounts when compared to the unfunded listings. Duration is slightly higher for the funded listings than unfunded ones. Finally, borrowers with funded listings have a lower debt-to-income ratio. Similarly, we also looked at the summary statistics of the loans that were paid back and not paid back (i.e., defaulted), see Table 2. Note that interest rate and debt-to-income ratio for loans that are not paid back is significantly higher than the ones that got paid. Also, while on average, the amount requested is higher for the defaulted listings, the number of bids is slightly lower.

Figure 1 shows that the probability of a listing being funded increases with the number of bids it receives. Notice that this probability increases sharply early on, but it flattens down as the number of bids increases.

Every listing on Prosper is assigned a Prosper Rating to analyze its level of risk. This rating represents an estimated average annualized loss rate range. The loss rate is based on the historical performance of borrowers on Prosper loans with similar characteristics and is determined by two scores: the first is the credit score, obtained from a credit reporting agency; the second is an in-house custom score, the Prosper Score, built on the Prosper population. The use of these two scores determines an estimated loss rate for

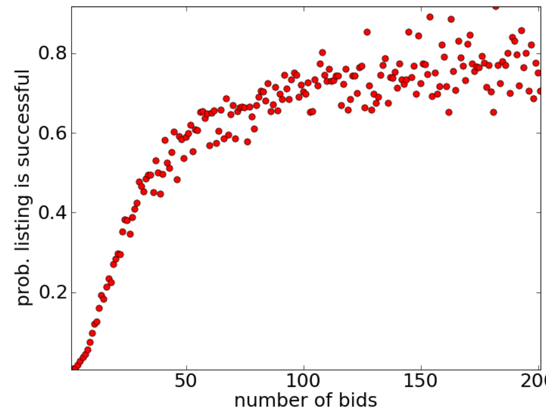


Figure 1: The probability of a listing being funded vs number of bids.

	T_n	T_{nb}	P_f	P_p	P_d	I
AA	11,454	1,048,816	34%	42%	9%	9.84%
A	13,747	928,896	28%	33%	16%	12.62%
B	20,776	1,038,617	26%	25%	20%	15.44%
C	36,655	909,821	20%	22%	25%	18.03%
D	50,577	617,965	15%	22%	29%	21.29%
E	59,888	271,069	9%	21%	40%	25.16%
HR	147,393	241,370	5%	15%	52%	25.04%

Table 3: Prosper Rating Statistics. See main text for column descriptions.

each listing, based on the historical performance of previous Prosper loans, which then determines the Prosper Rating.

Some summary statistics for each Prosper Rating is shown in Table 3. The first column shows the total number of listings, T_n. Note that there are much more listings created for lower Prosper Ratings. On the other hand, majority of the bids go to the higher Prosper Ratings according to the second column that shows the total number of bids, T_{nb}. The third column, where P_f stands for the percentage funded, shows that for lower Prosper Ratings, the success rate is also lower. Among the successful listings, percentage of the ones that are paid back is in the fourth column, P_p, while percentage of the defaulted ones is in column five, P_d. As expected, loans with higher Prosper Rating are more likely to be paid back and loans with lower Prosper Rating are more likely to be defaulted. However, lenders are still willing to bid for listings with higher risk since their interest rate is higher as shown in column five, where I stands for the average interest rate.

4. TEMPORAL PATTERNS AND MODEL

In this section, we investigate the dynamic features of bidding behavior of the lenders in Prosper. There are in total 923,457 registered members in the data set we study. However, most of the members do not have any activity in the time window of our data set. So we only focus on the users that are involved in the bidding activities, e.g. the users who have bid at least once or whose listings have received at least one bid. Among these users, borrowers are those who requested at least one loan, and lenders are those who made at least one bid. Based on this restriction, there are 136,080 users who are involved in the bidding activities. Among them, 48,824 are lenders who only bid on listings, 81,190 are borrowers who do not bid but create listings to ask for loans, and 6,066 are both lenders and borrowers.

As we have mentioned in the previous section, there are about 5 million bids in total. Among these 5 million bids, 3,281,070 bids

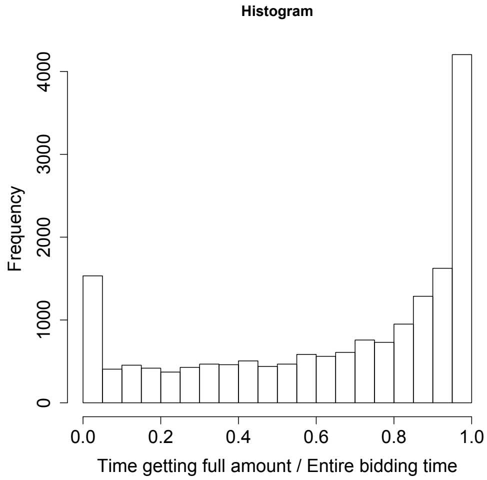


Figure 2: Distribution of time length of getting fully funded as a fraction of the entire time length of bidding of listings with competition.

go to successful listings, i.e. about 66% bids make the "correct choice". Moreover, we should notice that only 8% of the listings are successful. This is interesting as most of the lenders make very similar decisions and their bids go to only a small fraction of the entire listings, and most listings in this small fraction turn out to be successful in the end.

In this section, we mainly focus on the successful listings and the bids that go to them. Among these successful listings, about one third of them do not have *competition*. By *competition*, we mean that further bids are placed even after the total amount of all bids that have been placed in the listing equals the initial amount requested by the borrower. Thus, the entire bidding time of listings with competition can be split into two time intervals. First is the interval when the listing is accumulating bids. When the sum of the bid amounts is greater or equal to the requested amount, the listing is funded. As in the first part of the bidding, the goal is to collect the requested amount, all bids are in principle winning and there is no need for bidders to lower the interest rate. The second interval is the time period from the time when the initial amount requested is satisfied to the end of bidding, which we call as the "competition time". During this time period, the bids with highest interest rates are outbid by the incoming bids with lower interest rates, and the only way to win in the competition phase is to bid with lower and lower interest rates. On average, for the listings with competition, the time that they get fully funded is about 0.65 of the entire bidding time and 46% bids are received during this time, while the competition time is about 0.35 as long as the entire bidding time but 54% bids are received in this time period on average. Figure 2 shows the distribution of the fraction of the competition time to the entire auction duration of all the listings with competition. We observe that most of the listings with competition end soon after they receive enough money to be funded.

Since every listing has its own time span of auction, we use the following way to normalize the time of bidding. For the listings with competition the time scale ranges from 0 to 2, in which time 0 is when the listing receives the first bid, time 1 is when the listing gets fully funded (i.e., the first time when sum of bidded amounts is greater than the requested amount), and time 2 is the time that the listing receives its last bid. In this sense, time 1 to 2 is the competition time that is mentioned before. For the listings without competition, we use a time scale from 0 to 1 to represent the total duration of bidding.

4.1 Interest rates

First we investigate the change of interest rate over time. As described previously, when a user creates a listing i, she sets the maximum interest rate $\hat{I}_i$ she wants to pay. Throughout the bidding process, when the lenders bid, each lender indicates the minimum interest rate that she is willing to accept. At the end of a listing i, the final interest rate that the winning lenders get is: $\tilde{I}_i = \min I_k$, s.t., $\sum_{j,I_j<I_k} a_j = A_i$, where I_k is the minimum interest rate the lender would like to accept at the k-th bid of the listing i, a_j is the amount of money bid at the j-th bid, and A_i is the total amount of money the listing i requests.[1] Although the lenders who bid with lower interest rates have the advantage of winning the bids, they also possibly lower the final interest rate that the borrower will pay.

In order to study the change of interest rate of all the listings, we normalized each interest rate by the maximum interest rate $\hat{I}$ that is set by the borrower initially. For example, if the maximum interest rate the borrower is willing to pay for listing i is $\hat{I}_i$, then the normalized interest rate of j-th bid of the listing is $I_j/\hat{I}_i$.

In addition, we also examine the maximum interest rate that bidders could win with at every time t, $\tilde{I}_t$. More precisely, at any time $t \in [0, 2]$, $\tilde{I}_t$ is the maximum interest rate such that a new bid offering an interest rate not higher than $\tilde{I}_t$ would win. According to the definition, $\tilde{I}_t$ is equal to the maximum interest rate $\hat{I}$ the borrower would like to pay when $t \leq 1$. After $t > 1$, $\tilde{I}_t = \min I_k$, s.t., $\sum_{j,I_j<I_k} a_j = A_i$, and all bids k and j are before time t.

Figure 3 shows the change of average interest rate and the maximum winning interest rate of all listings with competition over time (the red solid line and the red dashed line), as well as the change of average interest rate of all listings without competition over time (the blue dashed line). As expected, the average interest rate remains around 1 for listings without competition as it is shown in Figure 3. This is because for listing with no competition, there is no incentive for bidder to lower the interest rate (i.e., they maximize the interest rate they can receive without having the risk of being outbid by others). Thus, it is flat and equal to the maximum winning interest rate $\tilde{I}_t$.

On the other hand, it is interesting to see that, at the time of 0-1, there is a large gap between the average interest I_t and the maximum interest rate $\tilde{I}_t$, and the average interest rate I_t is decreasing instead of being as flat as the maximum interest rate $\tilde{I}_t$. The big difference between I_t and $\tilde{I}_t$ at time 0-1 tells us that in order to win the bid, the lenders tend to lower their interest rates immediately after the listing starts, i.e., the actual competition starts soon after time 0 instead of after time 1. Because of this, we can see that the lenders are trying to balance between winning in competition and maximizing their profits since the very beginning. As a result of competition, the interest rates the lenders offer are always decreasing over time for the listings with competition, and the decrease is more significant during time 1-2 than the time period of 0-1. The curve of the average interest rate I_t and the maximum winning interest rate $\tilde{I}_t$ meet together when the time is close to 2. This tells us that when it is close to the end of bidding duration, the lenders have the advantage of having enough information about the auction and are able to make rational decisions – maximizing the interest rates they can earn while making sure that they can win the bids.

4.2 Probability of winning

As we have seen, the interest rates of listings change over time as

[1]Because of privacy issues of Prosper data, we do not know the interest rate of the winning bids. So in our study, we assume that all winning bids of a listing i are equal to the final interest rate, $\tilde{I}_i$. $\tilde{I}_i$ gives an upper bound of the interest rate of the winning bids.

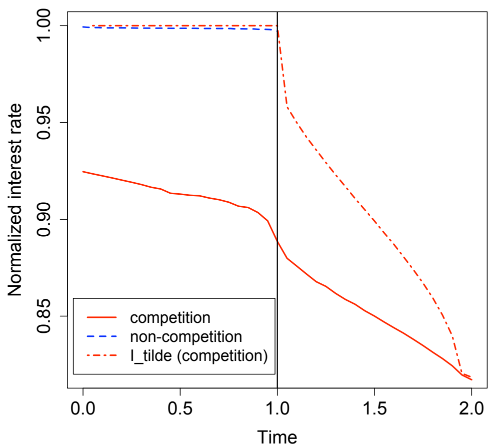


Figure 3: Average bidding interest rates of listings with competition (red solid line) and listings without competition (blue dashed line). $\tilde{I}$ is the maximum winning interest rate.

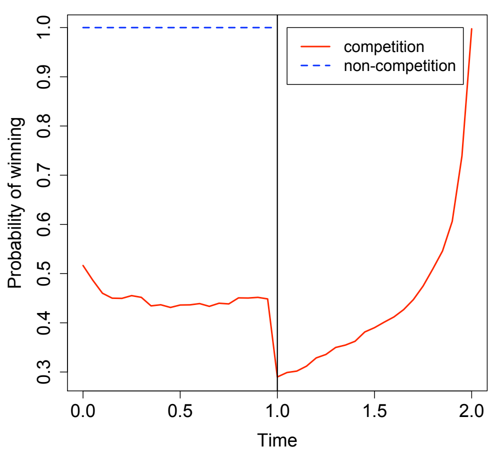


Figure 4: The probability of winning of bids over time. The red solid line is of the listings with competition, and the blue dashed line is of the listings without competition.

the lenders compete to win the bids. Next, we are going to investigate the probability of placing a winning bid as a function of time. For all bids, there are status labels showing whether the bids are successful (i.e., winning) or unsuccessful (i.e., outbid). For all the bids in the same time snapshot t, based on the status, we calculate the probability of winning. Figure 4 shows how the probability of winning changes as a function of time. Unlike the average interest rate of listings with competition, which monotonically decreases over time, the probability of winning does not consistently increase. It is interesting to see that the curve of probability of winning has very different behavior during the time interval 0-1 and 1-2: while the probability of winning remains almost constant during time 0-1, it exponentially increases during the time interval 1-2.

This tells us that under such a bidding mechanism, the lenders who bid close to the end of the bidding process would benefit from having most information about the bidding and thus win the bids with high probability. However, having more information does not always help. Lenders who bid before the listings get fully funded have advantage over those who bid immediately later than time 1 (i.e., after the listing collects enough bids to be funded). We also observe that there is a small peak when t is close to 0. These phenomena are due to the fact that if two bids have the same interest rate, then the early one wins (and thus the probability of winning is high at $t = 0$).

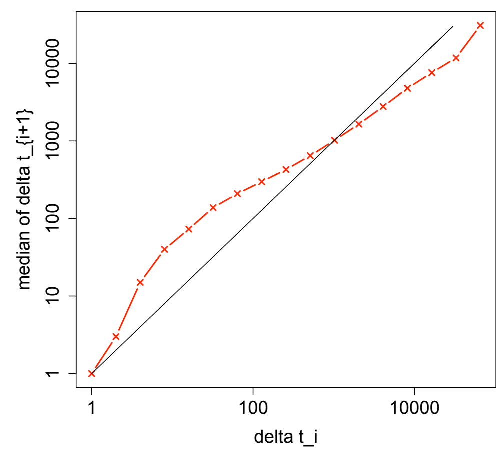


Figure 5: The time interval Δt_i of two consecutive bids b_i and b_{i+1} and the time interval Δt_{i+1} of the next two consecutive bids b_{i+1} and b_{i+2}.

4.3 Time of bidding

The question we investigate next is whether the lenders bid at a constant frequency over the entire time period when the listing is open, or whether there are specific time periods when lenders are more likely to bid.

In order to answer this question, we first look at the time interval Δt_i of two consecutive bids i and $i + 1$ and check if there is herding behavior during bidding. For all the values of Δt_i, we get the median value of all Δt_{i+1}. Figure 5 shows the relationship between Δt_i and the median of Δt_{i+1} after using logarithmic binning. From the figure, we see that Δt_i and Δt_{i+1} have a positively correlated relationship. If Δt_{i+1} is independent of Δt, we would expect a flat line. This positive correlation tells us that fast bids tend to be followed by fast bids, while slow bids tend to be followed by slow bids.

However, it is possible that the positive correlation is due to the fact that listings have different number of bids, and the bids are uniformly distributed in each listing. In order to distinguish the herding behavior from this case, we also plot the function $\Delta t_i = \Delta t_{i+1}$ (shown as the black straight line). The lenders would have bid uniformly over time in each listing if the curve overlaped with this straight line. If the curve is above the diagonal then it means the bids are decelerating – the time between bids $i + 1$ and $i + 2$ is longer than between i and $i + 1$. And similarly if it is below the line then it is accelerating – time between bids $i + 1$ and $i + 2$ is shorter than between i and $i + 1$. From Figure 5, we see that when Δt_i is less than around 1000 seconds (approximately 17 minutes), Δt_{i+1} is above the straight line, i.e., the bidding is decelerating; while Δt_i is bigger (more than around 1000 seconds), Δt_{i+1} is below the straight line, i.e., the bidding is accelerating and the next bid arrives sooner than the previous does. This suggests that there is herding behavior during the bidding process and the speed of bidding over time has ups and downs.

We further investigate if there is any special pattern of bidding of all listings with competition. We split the time period of each listing into 40 time snapshots (20 between time 0-1 and 20 between time 1-2 for listings with competition), and we count the number of bids that falls in each snapshot. Next, we sum up the numbers of bids of

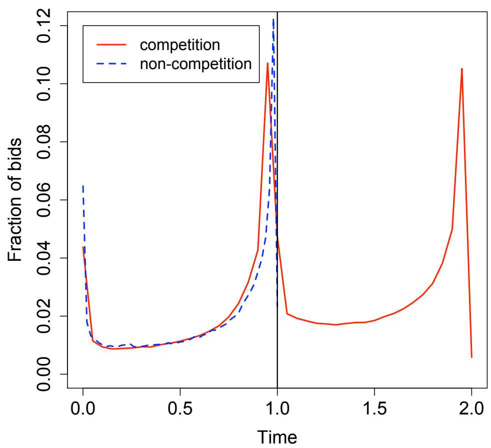


Figure 6: The fraction of bids over time. The red solid line is for the listings with competition, and the blue dashed line is for the listings without competition.

all listings at each time bin, and normalize them such that we get the fraction of bids over time as shown in Figure 6.

We see that there are two spikes at time around 0 and 1 for both successful listings with competition and successful listings without competition. For the listings with competition, there is one more spike at the end when t = 2. This suggests that for each listing, the bids it receives are not uniformly distributed over time. Instead, it is much more likely to receive bids around times 0, 1, and 2. So, lenders are more likely to bid at the beginning when the listing is fresh ($t = 0$), just before the time when the listing gets fully funded ($t = 1$) and just before the bidding ends ($t = 2$). This cannot be an effect of lenders coming from different time zones since all lenders are from US and the highest percentage of lenders are from CA. One possible explanation is the Prosper user interface, which allows potential lenders to sort listings according to percentage funded or time left. However, there are many other ways to sort listings such as title, category, amount requested, yield, and rating. In addition to that, there is an advanced search option where lenders can even use keywords to look for specific listings.

The three-spike bidding pattern is interesting, and now we attempt to explain it from a modeling perspective. What is the minimal set of dynamic factors that drives lenders bid in such an interesting and highly non-uniform pattern? Since this is an economic setting, we assume that bidders bid based on the expected profit, i.e., they bid with probability proportional to the expected utility they are getting. We hypothesize that there are three factors from the economic aspect that influence a lender's decision to bid on a given listing at any given time:

1. The probability that a lender believes a listing will get fully funded, $f(t)$. The "bandwagon effect" is the phenomenon that says people often do and believe things because they observe others do and believe the same thing. This effect is also called "herding instinct" [22]. According to the bandwagon effect, at every time t before the listing gets fully funded, a lender's belief that the listing will succeed $f(t)$ is based on the number of bids that have been accumulated by the listing from the beginning to time t, i.e., $f(t) \propto \int_0^t f(\tau)$, where N is the total number of lenders that are potentially able to bid. Let $g(t) = N \int_0^t f(\tau)$, then we have:
$$\frac{dg(t)}{dt} = f(t) \propto Ng(t), t \in [0, 1) \tag{1}$$
This means the more lenders bid previously, it is more likely that lenders will bid at current time. Solving the relation in Eq. 1, we get:
$$\begin{aligned} g(t) &= z_1 \alpha e^{Nt}, t \in [0, 1) && (2)\\ f(t) &= \frac{dg(t)}{dt} && (3)\\ &= z_1 N \alpha e^{Nt}, t \in [0, 1) && (4) \end{aligned}$$
where $\alpha = f(0)$, i.e., the number of bids a listing receives immediately after starting, and z_1 is a constant for normalization. On the other hand, when $t \in [1, 2]$, a listing is already funded, so $f(t) = 1$. Thus, we get that the probability that a lender believes a listing will get fully funded has exponential increase when $t \in [0, 1)$ and is a flat constant when $t \in [1, 2]$:
$$f(t) = \begin{cases} z_1 N \alpha e^{Nt}, & \text{when } 0 \le t < 1 \\ 1, & \text{when } 1 \le t \le 2. \end{cases} \tag{5}$$

2. The probability that a lender is able to win the bid, $p_w(t)$. We model $p_w(t)$ based on our measurements in Figure 4, which suggests that there are two separate regimes of behavior:
$$p_w(t) = \begin{cases} \beta, & \text{when } 0 \le t \le 1 \\ z_2(t-1)^s, & \text{when } 1 < t \le 2. \end{cases} \tag{6}$$
where $\beta \in (0, 1)$ is a constant, which describes the almost flat line of the probability of winning at time 0-1 in Figure 4. z_2 is a positive constant for normalization and $s > 1$. $z_2(t-1)^s$ describes the fast increase of the probability of winning at time 1-2 in Figure 4.

3. The estimated interest rate, $I_e(t)$. At time t, a lender estimates the current interest rate based on the average interest rate of other lenders at $t - \Delta t$. From Figure 3, we see that the average interest rate constantly decreases. So here, for the simplicity of this model, we assume that the estimated interest rate changes linearly with time t:
$$I_e(t) = a - bt, t \in [0, 2] \tag{7}$$
where a and b are constants, $a \in [0, 1]$ and $0 < b < a/2$.

In addition to the three economic aspects that lenders take into consideration when bidding, there is also a social aspect that influence lenders' bidding behavior. According to [23], people have preference for more recent news than old ones. To model this preference of users to new or fresh listings, we model this similarly as [17] and assume that the lenders' preference for new listings decrease over time in a polynomial function of t:
$$\delta(t) = ct^{-1}, t \in [0, 2] \tag{8}$$

In our model, a lender's bidding decision is based on Eq.s 5-8. I.e., over the entire bidding process, the probability for every lender to bid at a given time t is proportional to the product of three main factors; probability of funding, probability of winning and interest rate:
$$p_b(t) \propto f(t) p_w(t) I_e(t) \delta(t) \tag{9}$$

Figure 7 shows the simulation result of our model in Eq. 9. It has three spikes at time 0, 1, and 2. The first spike is a result of Eq. 8, which models lenders' preference for new listings. In Figure 6, we see that both the listings with competition and without competition have the first spike at time = 0. In fact, we also observe that this spike exists in almost all unsuccessful listings as well. This fact

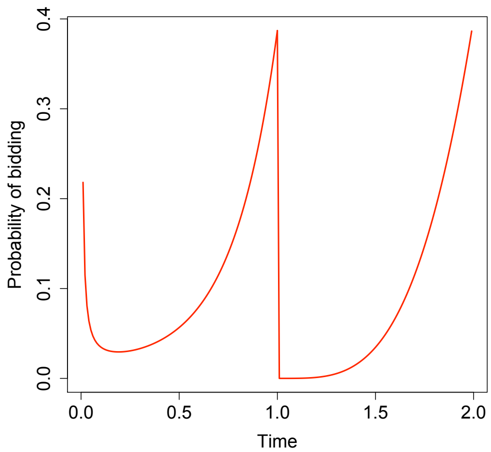


Figure 7: Bidding probability over time $p_b(t)$ simulated by our model.

further supports our modeling decision that the first spike is not a result of any economic factor, but rather a result of a social effect. The second spike at time 1 is a result of an economic factor, i.e., lenders' belief that a listing will get fully funded is affected by others' behavior. Thus, the second spike at $t = 1$ means that bidders who think the listing will get paid back want to bid early to maximize their profit (i.e., win with a bid of high interest rate). We also observe the second spike of both listings with competition and without competition in Figure 6, but we do not see this spike in the listings that do not success. The third spike in the listings with competition is stimulated by the probability of winning in Figure 4, and we can also see that this growth only exists in the listings with competition. Lenders who care more about the probability of winning than the profit are more likely to bid at this time and thus generate this spike.

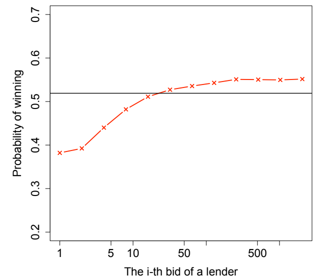
As we discussed in Section 3, Prosper assigns different credit ratings from AA to HR to borrowers. In addition to the empirical results we show in Figure 3, 4 and 6, we also examine the bidding dynamics of listings whose borrowers belong to different groups of credit ratings. We find that listings grouped by different credit ratings have similar dynamic curves as we see in Figure 3, 4 and 6. This means that all listings behave qualitatively similar regardless of what credit score they are from.

4.4 Performance of individual lenders

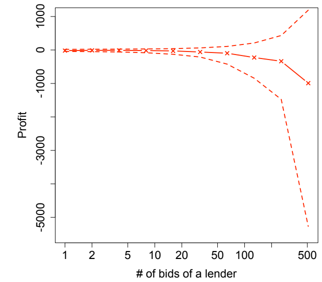
In the previous parts of the section, we focused on the aggregated behavior of all lenders in listings. In this part, we investigate the performance of individual lenders. There are two ways to measure the performance of individual lenders. One is the lenders' probability of winning the bid, and the other is the profit they make.

We have also calculated how much net profit each lender at Prosper made so far. We only took into account the loans that have been either paid back or defaulted. For the winning bids, based on the amount bid by each lender and the final interest rates of the loans, we can get how much money a lender made or lost from each bid he made. While 11,182 lenders are even, i.e. they neither lost money nor made profit, 23,596 lenders lost money and 9,087 of them made profit.

The first question we are going to answer is whether more experienced lenders have better performance, i.e., they are more likely to win the bids and make more profit. Figure 8(a) shows the relationship between the i-th bid of all lenders and the average probability of winning the bid. The black line shows the overall winning probability, which is slightly above 0.5. As we see in Figure 8(a), the probability of winning grows as the lenders get more experienced. When lenders bid over 50 times, they have better chances to win than average.



(a)



(b)

Figure 8: (a) The average probability of winning the bids versus the i-th bid of a lender. (b) The profit the lenders make versus the number of bids they have. The red solid line is the median value and the red dashed lines show the upper and lower 5% values.

	Time 0	Time 1	Time 2
Prob. of winning	0.0216*	-0.0766***	0.2158***
Profit	-0.0102*	-0.0820***	0.0280*

Table 4: Correlation between bidding behavior and performance of lenders who have bid > 200 times. * means the p-value > 0.05, and * means the p-value < 0.001.**

On the other hand, we are also interested in the number of bids and the profit a lender makes. Again we would expect that more experienced lenders with many bids make larger profits while inexperienced users make losses. However, Figure 8(b) shows that there is a slightly negative correlation between the profit a lender makes and her total number of bids. In addition to the overall correlation, we also see that the variance grows larger as the number of bids grows.

In order to further understand how the bidding behavior is related to lenders' performance, we look at the time distribution of all bids of every single lender who has more than 200 bids in successful listings with competition. The hypothesis here is to check whether experienced users tend to bid at particular points in the bidding process. First we compute the density of bids around time 0, 1, and 2 of every lender. Then we correlate the density of bids around time 0 (or time 1 or 2) with the probability of winning and profit of every lender. Table 4 shows the results of correlation. We see that lenders who tend to bid around time 1 are less likely to win the listings, while lenders who tend to bid around time 2 are more likely to win the bids. This is consistent with the Figure 8(a), as the probability of winning around time 0 is about equal to the overall probability of winning. However, bidding around time 1 is lower than average and bidding around time 2 is much higher. From Table 4, we also see that for a lender, the tendency to bid around time 0 or 2 is neutrally correlated with the profit she can make; while bidding around time 1 is negatively correlated.

Moreover, Prosper suggests lenders to bid on more listings with small amounts of money in order to minimize the risk. In order to verify the strategy of decentralization, we count the number of distinct successful listings each lender j bid n_j^d. We divide n_j^d by the total number of bids of each lender n_j. The value of n_j^d/n_j of lender j indicates how decentralized the lender bids. Again, we correlate this value with the profit the lenders make, and get a weak positive correlation 0.02 (with p-value < 0.001). This means lenders slightly benefit from the strategy of decentralizing their bids.

5. PREDICTING THE LOAN SUCCESS

In Prosper, listings for which at least 100% of the requested amount is collected, are considered "fundable" (successful) and they translate into an active loan. However, listings which do not reach full funding are considered unsuccessful ("not fundable") and no loan is created. Out of the loans that are funded, some are repaid on time and others are cancelled or their borrowers default on them.

In this section, we examine a simple model that predicts whether a listing is going to be funded or not, and whether it will be paid back or not. A similar study is conducted at [11] and [20], where the authors focus on borrower and listing attributes. Their goal is to provide a ranking of the relative importance of various fundability determinants, rather than providing a predictive model. However, our goal here is different as we do not just want to make our predictions based on some large number of features but are instead interested in modeling how the temporal dynamics of bidding behavior predicts the loan outcome (funded vs. not funded and paid vs. not paid). Thus we are interested in how much signal is in "how the market feels" as opposed to traditional features such as credit grade or debt-to-income ratio.

We started our analysis by looking at the time series history of loan listings. In other words, we examine the progression of the total amount bid on a given loan as a function of time. We used a time scale from 0 to 1, in which time 0 is when the listing receives the first bid and time 1 is when it gets the last bid. Let A_i be the total amount bid for listing i and $\sum_{j \leq k} a_j = A_k$, where a_j is the amount of money bid at the j-th bid, so A_k is the total amount of money bid till the k-th bid. For each listing, we looked at $Y_R = \frac{A_k}{A_i}$ as a function of time. Figure 9 shows the four main types of curves we observed. This observation led us to the hypothesis that the total amount bid on a given listing follows a sigmoid curve as a function of time. As a result, we fit a sigmoid (logistic) curve to each listing time series, defined by

$$y(t) = \frac{1}{1 - e^{-q(t-\phi)}},$$

and we used least squares to find the optimal q and ϕ. Parameter q controls how quickly the function rises while ϕ controls the time (x-value) at which the rise occurs.

For each listing's fit, we calculated the R-squared error. The average R-squared error is 0.9, which shows that overall we do a good job of fitting the data. This is not our main goal, however. We wish instead to use the shape parameters, q and ϕ of a listing's bid history to predict whether or not this listing will be funded and paid back.

Some examples of our fitting can be seen in Figure 10. i^{th} dot depicts the total fraction of collected money at the time of i^{th} bid of that particular listing and the smooth curves are the fitted logistic curves. While q is a measure of the steepness of the curve, ϕ tells us where the inflection point of the sigmoid curve is located. Mainly, all the listings fall into one of the four curve types as shown in Figure 9. For low q and high ϕ, the curve has a less steep sigmoidal shape. For high q and high ϕ,the curve has an exponential shape. For low q and low ϕ, the curve has diminishing returns shape and for high q and low ϕ, the curve has a steep sigmoidal shape.

Figure 11 shows a plot of q versus ϕ both for funded (purple triangles) and non-funded (blue circles) listings. The two classes are mostly distinguishable, especially in the middle range of values for both q and ϕ. This is similar for loans that have been paid back

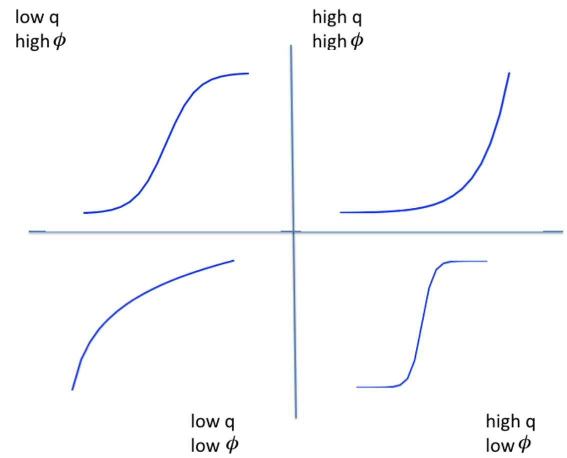


Figure 9: Main curve types that were observed when we plot total fraction of collected money as a function of time for each listing.

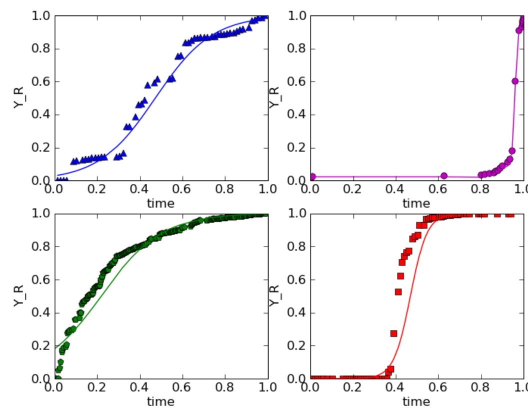


Figure 10: Real instances of what Figure 9 illustrates. Each dot is a bid of that particular listing, smooth curves are the fitted logistic curves.

(green triangles) and those that have not (red circles) as shown in Figure 12.

In order to verify the importance of q and ϕ in predicting the success of a listing, we constructed a logistic regression prediction model that uses these two quantities as features. As discussed in Section 3, funded listings have significantly larger number of bids than the non-funded ones. So, we also included the total number of bids, N_b as a parameter which helps the model. Table 5 shows the summary of the regression model that predicts the success of a listing. According to the table of coefficients, both q and ϕ are significant predictors of success of a listing. For every one unit change in q, the log odds of success (versus non-success) increases by 0.063. For a one unit increase in ϕ the log odds of a listing being successful decreases by 0.7162. In other words, the higher the steepness of the curve, the more likely a listing will be funded and the sooner the curve spikes (negative ϕ coefficient) the better. So, observing a steep sigmoidal curve for the progression of the total amount bid for a listing is a good sign of its success.

We used cross validation to understand how well the regression model works, i.e., we split the available data into five buckets, trained our regression model on four of them, tested the accuracy on the remaining one and repeated this procedure for each test bucket.

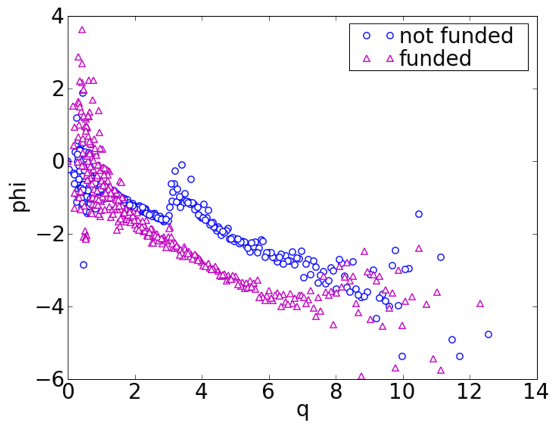


Figure 11: q vs ϕ for funded and non-funded listings.

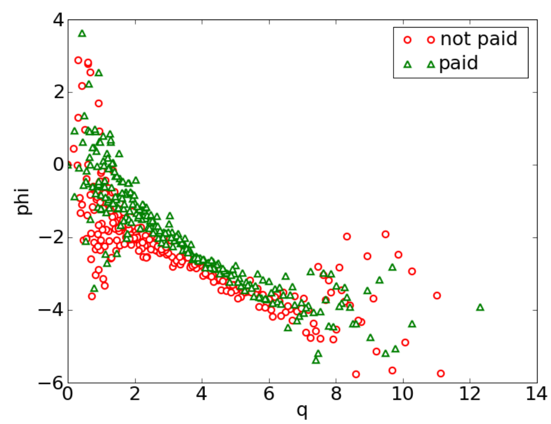


Figure 12: q vs ϕ for paid back and not paid back listings.

As a result of the cross validation, the prediction accuracy of the above logistic model is 95%. Since only 8% of the listings are funded we repeated the same analysis, however, this time making sure that both the training and test sets have a balanced amount of positive and negative examples, i.e. we under-sample the non-funded listings. Using cross validation as explained before, gave us a prediction accuracy of 87% while simple random guessing would give 50%. We also investigated the performance of the model with incomplete data by only using half of the bids for each listing to fit a sigmoid curve to its time series. We observed that shape parameters are still good predictors of success since the prediction accuracy only dropped to 85% from 87%. In [20] and [11], authors study the relative importance of various fundability determinants, such as borrower characteristics and loan attributes. In order to see how well the new features we introduced complement the ones proposed in earlier studies, next, we combined all these features. We picked the following standard features as they are the most affective ones; maximum rate borrower is willing to accept b_r, debt-to-income ratio d_r , total amount requested T, whether or not borrower is a home owner h and listing description length d_l. We will call these five set of features as "baseline features". The prediction accuracy of the logistic regression with the baseline features is 67% while the accuracy increases to 92% if we combine them with the "bid features"; ϕ, q and N_b.

We also constructed a logistic regression model for each Prosper credit rating category with the same features used for training the

	Estimate	Std. Error	z value	Pr(> $\|z\|$)
(Intercept)	-3.572	0.0197	-180.716	0.0000
q	0.0636	0.0064	9.916	0.0000
ϕ	-0.7162	0.0082	-86.62	0.0000
N_b	0.012	0.0001	-89.766	0.0000

Table 5: Regression model for predicting whether the listing will get funded or not.

RATING	BASELINE F.	ϕ and q	BID F.	ALL
AA	65%	60%	75%	86%
A	66%	70%	83%	90%
B	65%	73%	85%	92%
C	66%	75%	88%	95%
D	66%	81%	91%	96%
E	72%	86%	93%	96%
HR	71%	89%	91%	94%
Average	67%	76%	87%	92%

Table 6: Prediction accuracy for listing success (funded vs. not funded) per Prosper rating.

classifier. Table 6 shows the prediction accuracies of the logistic regression models that were constructed by using all available data for each Prosper rating. Again, for each category the data is highly skewed, i.e., the majority of the listings are not funded (see the column that shows the percentage funded, Pf, at Table 3 in Section 3). Thus, we sampled a balanced amount of funded and non-funded listings for each rating before applying the regression analysis.

The first column of Table 6 shows the prediction accuracies when only the baseline features are used for the regression while the second column is for the model that only includes the shape parameters, ϕ and q. Using only baseline features performs worse than the shape parameters. ϕ and q only model the temporal progression of the time series but not the number of bids itself so, in third column, we added total number of bids to shape parameters, which gives better results. However, it is important to note that only using the shape parameters also heavily improves over random guessing, which would give 50% accuracy. Overall, we get the highest improvement for the hardest cases, such as the lowest Prosper Rate HR and E while the smallest improvement is for AA. The last column lists the accuracy of the model when the baseline and bid features are combined, which outperforms others. These results show that there is information to be gained from "how the market feels" that is not present in the set of features constructed from standard factors such as homeownership or debt-to-income ratio.

Next, we conducted the same analysis to predict whether a listing will be paid-back or not. This means that we aim to predict whether the listing will get paid back by the borrower solely based on the temporal dynamics of bidding for that listing. This time the data set was filtered to contain only the listings that got funded. Since some of the loans were still ongoing, i.e., their repayment was not over, we only picked the ones that were paid-back or defaulted. We ran logistic regression for different attribute sets. In addition to shape parameters, adding the following features of a listing gave the best results; borrower rate b_r, total amount requested T and whether there has been competition or not c. The summary of this regression model is in Table 7. By looking at the coefficients of the logistic regression, we can say that the shape parameters, ϕ and q, are both positively correlated with its being paid back or not. Similar to funding success, the higher the steepness of the curve, the more likely a listing will be paid. However, this time the later the curve spikes the better, which is a signal of competition. Table 7 shows that the existence of competition have positive impact on a loan's being paid-back or not while the maximum rate that the borrower accepts and the amount requested have a negative impact.

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
(Intercept)	2.567	9.145	28.066	0.0000
q	8.575	1.722	4.980	0.0000
ϕ	1.219	1.807	6.749	0.0000
b_r	-1.260	3.110	-40.506	0.0000
T	-5.227	3.554	-14.710	0.0000
c	5.179	4.264	12.144	0.0000

Table 7: Regression model for predicting whether the listing will get paid-back or not.

RATING	BORROWER F.	ϕ and q	LOAN F.	ALL
AA	70%	85%	87%	78%
A	62%	69%	72%	73%
B	63%	58%	67%	70%
C	59%	57%	64%	64%
D	57%	60%	60%	61%
E	55%	61%	62%	61%
HR	77%	78%	80%	77%
Average	63%	67%	70%	69%

Table 8: Prediction accuracy of being paid-back or not per Prosper rating.

When we tested our model through cross validation, the prediction accuracy was 70% this time. On the other hand, using only the shape parameters, ϕ and q, gave 67% accuracy. If we add some standard features to the model such as Prosper rate, debt-to-income ratio d_r , whether or not borrower is a home owner h and listing description length d_l, the accuracy only increases to 72% from 70%. The prediction accuracy of the regression model for each rating with balanced data is in Table 8. Only using borrower features; d_r, h, d_l improves over random guessing, which would give 50% accuracy, see first column. However, using only shape parameters performs better than borrower features, see second column. Our main model, third column, i.e. adding some loan features, b_r, T and c to shape parameters gives the best results even better than combining loan features with borrower features, see the fourth column. Note that prediction accuracy does not decrease monotonically with Prosper Rating. We get the highest improvement for the Prosper Ratings HR and AA, while the smallest improvement is for D. So, predicting a loan's being paid back or not is harder for the mid ranges of Prosper Ratings such as C and D.

To the best of our knowledge, this is the first study that also focus on the prediction of the performance of loans in addition to fundability success. We showed that only exploring the temporal dynamics of a listing instead of the borrower characteristics is a predictor of timely payments.

6. CONCLUSION

In this paper, we studied dynamics of bidding mechanism in a peer-to-peer lending marketplace, Prosper. We investigated the factors that affect the bidding process throughout a listing's lifetime.

We observed that bids for a single listing do not occur uniformly over time. We concluded that this is a result of lender's taking into account three factors while bidding on a listing: interest rate, probability of being amongst the winning lenders, and overall probability of a listing being successful. These factors change over the lifetime of a listing, thus explaining the non-uniform distribution of bids over time. We also examined the performance of individual lenders and observed that the profits the lenders make are tied with their bidding preferences.

Finally, we built a logistic regression model to predict listing's fundability success, as well as a model used to predict the probability of a loan's being successfully paid back. We showed that the listing bid trajectory plays an important role on both of these models and only based on the temporal progression of the bidding behavior, we can well estimate listing success.

Acknowledgements Research was in-part supported by W.R. and S.H. Kimball Stanford Graduate Fellowship, NSF Award 0835614, NSF CNS-1010921, NSF IIS-1016909, AFRL FA8650-10-C-7058, A. Yu & M. Bechmann Foundation, IBM, Lightspeed, Microsoft and Yahoo.

7. REFERENCES

[1] K. Asdemir. Bidding patterns in search engine auctions. *In EC'06*, 2006.
[2] J. Auerbach, J. Galenson, and M. Sundararajan. An empirical analysis of return on investment maximization in sponsored search auctions. *In ADKDD'08*, 2008.
[3] S. Berger and F. Gleisner. Emergence of financial intermediaries on electronic markets: The case of online p2p lending. *Business Research*, 2(1), 39-65 2009.
[4] N. Chen, A. Ghosh, and N. Lambert. Social lending. *In EC'09*, 2009.
[5] T. Dietrich, D. Seese, and S. Chalup. Classification of ebay bidding characteristics. *IADIS International Journal on WWW/Internet*, 4:111–125, 2006.
[6] B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. *In EC'05*, 2005.
[7] S. Freedman and G. Jin. Do social networks solve information problems for peer-to-peer lending? Evidence from prosper.com. *Working Paper*, 2008.
[8] S. Freedman and G. Jin. Dynamic learning and selection: the early years of prosper.com. *Working Paper*, June 2008.
[9] K. Ganchev, A. Kulesza, J. Tan, R. Gabbard, Q. Liu, and M. Kearns. Empirical price modeling for sponsored search. *Workshop on Sponsored Search Auctions, WWW'07*, 2007.
[10] Gartner. http://www.gartner.com/it/page.jsp?id=1272313.
[11] M. Herzenstein, R. Andrews, U. Dholakia, and E. Lyanders. The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities. *papers.ssrn.com*, June 2008.
[12] M. Hulme and C. Wright. Internet based social lending: Past, present and future. *Social Futures Observatory*, 2006.
[13] R. Iyer, A. Khwaja, E. Luttmer, and S. Kelly. Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending? *HKS Faculty Research Working Paper Series RWP09-031*, 2009.
[14] Kiva. http://www.kiva.org/.
[15] M. Klafft. Online peer-to-peer lending: A lender's perspective. *In EEE'08*.
[16] LendingClub. http://www.lendingclub.com/.
[17] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. *In KDD '09*, 2009.
[18] D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves. Pennies from ebay: the determinants of price in online auctions. *Journal of Industrial Economics*, 2007.
[19] Prosper. http://www.prosper.com/.
[20] J. Ryan, K. Reuk, and C. Wang. To fund or not to fund: Determinants of loan fundability in the prosper.com marketplace. *Stanford Graduate School of Business*, 2007.
[21] H. Shah, N. Joshi, and P. Wurman. Mining for bidding strategies on ebay. *In WEBKDD'02*, 2002.
[22] Wikipedia. http://en.wikipedia.org/wiki/bandwagon_effect.
[23] F. Wu and B. Huberman. Novelty and collective attention. *PNAS*, 104:17599–17601, 2007.