



Hidden Indicators of Collective Intelligence in Crowdfunding

Emőke-Ágnes Horvát

Northwestern University

Evanston, IL, USA

a-horvat@northwestern.edu

Jayaram Uparna

Indian Institute of Management Udaipur

Udaipur, India

jayaram.uparna@iimu.ac.in

Henry K. Dambanemuya

Northwestern University

Evanston, IL, USA

hdambane@u.northwestern.edu

Brian Uzzi

Northwestern University

Evanston, IL, USA

uzzi@kellogg.northwestern.edu

ABSTRACT

Extensive literature argues that crowds possess essential collective intelligence benefits that allow superior decision-making by untrained individuals working in low-information environments. Classic wisdom of crowds theory is based on evidence gathered from studying large groups of diverse and independent decision-makers. Yet, most human decisions are reached in online settings of interconnected like-minded people that challenge these criteria. This observation raises a key question: Are there surprising expressions of collective intelligence online? Here, we explore whether crowds furnish collective intelligence benefits in crowdfunding systems. Crowdfunding has grown and diversified quickly over the past decade, expanding from funding aspirant creative works and supplying pro-social donations to enabling large citizen-funded urban projects and providing commercial interest-based unsecured loans. Using nearly 10 million loan contributions from a market-dominant lending platform, we find evidence for collective intelligence indicators in crowdfunding. Our results, which are based on a two-stage Heckman selection model, indicate that opinion diversity and the speed at which funds are contributed predict who gets funded and who repays, even after accounting for traditional measures of creditworthiness. Moreover, crowds work consistently well in correctly assessing the outcome of high-risk projects. Finally, diversity and speed serve as early warning signals when inferring fundraising based solely on the initial part of the campaign. Our findings broaden the field of crowd-aware system design and inform discussions about the augmentation of traditional financing systems with tech innovations.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; **Empirical studies in HCI**.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW '23, April 30–May 04, 2023, Austin, TX, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9416-1/23/04.

<https://doi.org/10.1145/3543507.3583414>

KEYWORDS

crowds, collective decision-making, opinion diversity, aggregation speed, early warning signals

ACM Reference Format:

Emőke-Ágnes Horvát, Henry K. Dambanemuya, Jayaram Uparna, and Brian Uzzi. 2023. Hidden Indicators of Collective Intelligence in Crowdfunding. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*, April 30–May 04, 2023, Austin, TX, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3543507.3583414>

1 INTRODUCTION

An impactful discovery in decision-making research has been “crowd wisdom,” or the ability of the crowd to make better decisions than individuals make [7, 48, 54, 74, 103, 104]. Recent research has begun exploring the important question of whether the wisdom of crowds operates in complex human decision-making situations [13, 34, 51, 75, 80, 84, 86, 89]. Currently, the wisdom of crowds studies show that crowds can produce better estimates of *unknown knowns*—decisions where true answers are known a priori but not by the crowd. The original case of unknown knowns dating back to the early twentieth century involved guessing the weight of a quartered ox at an English countryside fair. The closest guesser would win the quartered ox. Only the butcher knew the ox’s true weight. The remarkable discovery was that the crowd’s estimate of the ox’s weight was better than any person’s in the crowd, including people with expertise in cattle ranching and butchery. The crowd’s guessed weight, measured as the average of all guesses, was off by just 11 pounds (the true weight was 1109 pounds) [46]. The power of crowds to guess unknown knowns has been replicated in diverse contexts [92, 95], including estimating the number of jellybeans in a jar [12].

In contrast to unknown knowns, complex decisions involve estimating *unknown unknowns*, where the true value becomes known in the future instead of at the time of the guess¹. Complex decisions are widespread and include judgments about investments, patients’ reactions to therapies, the impact of climate on migration,

¹We argue that unknown unknowns are a crucial challenge in crowd decision-making, just as they are a weak spot in machine learning. In the case of machine learning, unknown unknowns represent examples where a model is confident about its answer but is wrong, as opposed to signaling that it is unsure of the answer [64, 91, 109]. Work in human computation demonstrates that humans can identify unknown unknowns in machine learning models, recognizing rare but important errors [9]. Motivated by this evidence, we go one step further and use the concept of unknown unknowns to distinguish the toughest problems for collective intelligence that do not have a clear true value at the time of estimation. Specifically, our paper tests the ability of the crowd to provide superior decisions in complex real-world unknown unknowns.

insurance risks, mission failure, or the replicability of scientific findings [19, 26, 39]. Typically, complex decision-making is left to specially trained or experienced experts [96]. However, there is increasing evidence that expert decision-makers are more inaccurate and biased than was previously assumed [15, 80, 96]. Celebrated experts are barely more accurate than informed non-experts and, depending on the circumstances, as low as 12% of the outcomes they predict actualize [15, 80, 96]. This suggests that the exploration of crowd wisdom for complex decision-making has both theoretical and practical importance [80]. Together, the limitations of expertise and the remarkable estimation accuracy of crowds raise new questions about whether crowd wisdom can provide decision-making support for complex decisions.

Web systems enable broad tests of crowd wisdom's potential application in complex decision-making. A growing area of crowd decision-making includes crowdfunding decisions, which is an important scientific study on its own [47, 79, 94]. Over the past years, crowdfunding has grown rapidly in popularity, with some projecting that the global crowdfunding market will reach USD 28.9 billion by 2028 [106]. Crowdfunding could be especially critical in sectors under-served by traditional financing institutions [8, 73]. Our analysis studies investment decision-making on one of the first crowdfunding platforms for personal loans in the U.S. Crowdfunding involves untrained novices making complex decisions about loaning their funds. Funds are loaned to anonymous borrowers who promise to repay the loan's principle with interest but who pledge no collateral in the event of default, meaning that the crowd bears the entire risk associated with financing. In crowdfunding, loan decisions rely on collective intelligence because the crowd agrees without the aid of a leader or formal organizational rules to fund or not fund the loan.

Collective intelligence is linked to a crowd's diversity of opinions and the speed at which opinions are aggregated. When estimates are diverse, there is a greater likelihood that they will be equivalently dispersed above and below the true value, thereby averaging out estimation errors [55, 83]. When opinions arrive quickly, one after the other, the increased aggregation speed enables novel or specialized knowledge to be incorporated early into the crowd's decisions, thus avoiding bias toward lowest common denominator information [71, 84, 95]. How these factors operate in complex decision situations involving unknown unknowns is unclear because unknown unknowns have no a priori true value, and decision-makers are predisposed to social influence [28]. Experiments indicate that when individuals find it difficult to evaluate an object's true quality, they are susceptible to herding, which restricts the diversity of opinions [88]. Similarly, fast aggregation can emphasize the first few estimates of quality, which can lead to collective decisions that are biased [14, 41, 81].

We hypothesize that diversity and speed positively relate to prediction accuracy in real-world decisions but with a contingency. Because real-world decisions are complex and concern unknown unknowns, the mechanisms that generate collective intelligence need to be stronger for such decisions than they would need to be when the crowd is estimating unknown knowns. This suggests that the collective intelligence of the crowd for complex decisions with unknown unknowns requires both a high level of diversity and high speed for accurate collective decision-making. If either or

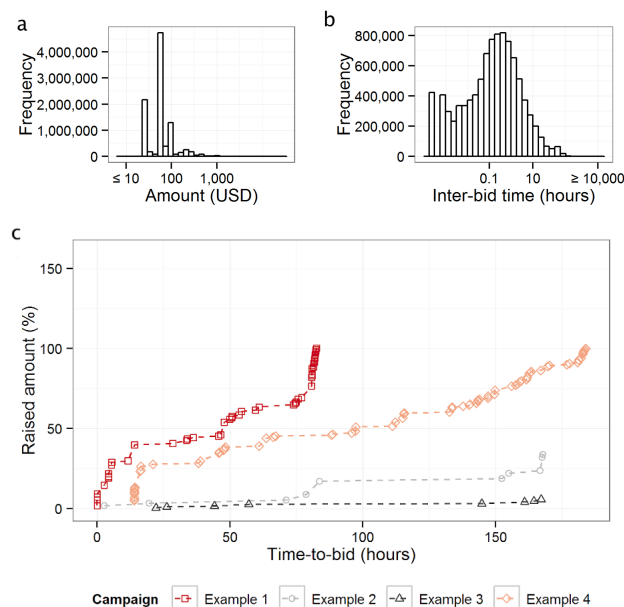


Figure 1: Histograms of (a) contribution amounts and (b) inter-bid times. Evolution of campaigns on the lending platform (c): the raised capital as the percentage of the target amount is shown in function of time for successful and unsuccessful campaigns (depicted in tones of red and gray, respectively).

both of these collective intelligence indicators are low, the crowd's decisions will be no more accurate than what is expected by the simple base rate. Specifically, we test whether crowds make accurate funding decisions as predicted by the diversity of opinion and the speed at which people collectively bid on a loan. A related theoretical and practical test is whether a few people reach the same decision as large groups in complex real-world estimation settings [82]. Due to their size, large crowds are intuitively likely to be more diverse and have more novel information. This raises the question of whether diversity and speed based on the first few contributors only (i.e., a small crowd) are associated with the funding decisions that large crowds made.

This paper thus puts to the test two non-trivial collective intelligence indicators (diversity and speed), finding empirically that they may benefit lenders and borrowers alike by promoting strategies for timely reinvesting and campaign retraction to avoid credit deterioration. Conceptually, our work identifies a compelling example of collective intelligence working better in practice than theories have predicted. In the human computation community, our findings call for concerted efforts to develop new theories of the wisdom of crowds that are inspired by parallel research in Web-enabled collective intelligence systems.

2 RELATED WORK

The idea of human computation as a means to solving computational problems in crowdsourcing applications has attracted substantial research attention from diverse fields, including artificial

intelligence and human-computer interaction (HCI) [87]. Relatedly, the notion that large groups of loosely organized people can arrive at solutions superior to those invented by any individual in the group has gained prominence in human computation research and allowed for several real-world applications of collective intelligence [74]. While human computation is related to, but not synonymous with, collective intelligence, most human computation systems comprise “groups of individuals doing things collectively that seem intelligent” [74] rather than individuals working in isolation. This recognition raises the question of how to optimally harnessing collective intelligence in human computation systems.

Despite rapidly growing research that taps into the collective intelligence accumulated on the Web, there are substantial challenges with harnessing the wisdom of crowds [16, 38, 59, 63, 85]. For example, effectively implementing crowdsourcing systems that rely on collective intelligence is challenging due to difficulties in ensuring that crowd members pay sufficient attention to the task, which impacts the quality of their responses [59, 85]. Enabling effective coordination among crowd members in crowdsourcing applications can likewise prove challenging in the sense that applications must dynamically adapt the system’s behavior to the quantity, quality, and reliability of users’ responses [16]. As another example, providing appealing and appropriate incentives can also be difficult as users’ motivations may not always be well-aligned with the goals of the collective task [59].

Given these common challenges in crowdsourcing, crowdfunding platforms provide an ideal setting for scrutinizing the emergence of collective intelligence. Crowdfunding entails several individual decisions that yield collective outcomes. This process lends itself to thorough empirical study. Access to comprehensive data on successful and failed projects enables us to understand the information available to the crowd that represents the basis for crowd members’ decision-making. With these data, one can track to the minute individual bids and the outcome of the project. Additionally, crowdfunding platforms represent a high-stakes decision-making setting where people risk financial losses. This personal investment strengthens users’ attention to and investment in the task.

There is ample evidence that the quality of the project description [5, 65, 78], project updates [105], the use of media content such as photos and videos [66, 105], and promotional activities on social media [43, 69, 108] are important factors in securing the success of a fundraising effort. Some research indicates that the type of project also matters because people tend to support projects that reflect their interests or values [70, 79]. Additionally, projects that request large amounts of money are more likely to fail to reach their fundraising goal than small asks [32, 33, 79]. Several studies also suggest that the reputation of the project creators [30], their social capital [27, 31, 56, 58, 79], and their ability to attract funders early in the campaign [3, 20, 24, 31, 93, 100] impact whether they can secure funding or not.

Few studies have examined the probability that projects deliver. The majority of these studies focus on lending and rely on attributes of the borrowers (e.g., annual income, credit history, housing situation, indebtedness, etc.) or loans (e.g., loan size and interest rate) [23, 42, 90] to predict loan default. Prior research that focuses on loan characteristics also shows that bad credit and high interest rates are associated with a greater likelihood of loan default in

crowdfunding as well [23, 42, 60]. According to this literature, a borrower’s credit information is regarded as the most informative predictor of loan default, even in online crowdfunding settings.

Previous studies have also developed crowd-based indicators of crowdfunding success. These indicators describe the overall number of contributions [24], the contribution amounts [25, 62] and timings [31, 93], and the amounts raised over a certain time frame [20]. Data from donation, equity, and lending-based crowdfunding platforms show that crowd signals similar to opinion diversity and aggregation speed are the most important in predicting who receives funding [37]. Our paper builds on these studies to establish empirically whether such signals are genuinely collective intelligence indicators that predict repayment or if they simply forecast the engagement or herding needed for successfully raising funds. This is critical because prior research shows that while herding [25, 107] and network effects [3, 99] facilitate fundraising, they can also lead to collective decisions that converge on sub-optimal lending outcomes in the long-term [107].

Our work differs from previous research in that we predict both funding and default, not based on project-related factors, social characteristics, or the amount of money bid as a function of time, but on two universal variables, the crowd’s opinion diversity and the speed at which the crowd contributes funds. Moreover, we also test the predictive power of these indicators based on small crowds. Additionally, we explicitly focus on whether crowds are relatively “smart” in predicting loan funding and default. Correctly predicting repayment for low-risk loans may be relatively easy, because the background rate of repayment for these loans is already high. One measure of a smart crowd is that it can predict the outcome of risky loans better than expected by the background rate. Thus, we model new collective intelligence problems in novel ways, using universal variables. In this way, we contribute to the expanding body of research that aims to harness the wisdom of crowds in online settings [34, 36, 92].

3 CROWDFUNDING PLATFORM & DATA

We use a data set that contains verified funding and repayment information, as well as fine-grained information on individual lender contributions. We analyze all financial transactions from Prosper.com between January 2006 and March 2012. Over this six-year period, 29,494 borrowers posted 236,409 loan requests, on which 65,756 lenders made 9,702,242 bids. 179,445 campaigns that did not receive bids were omitted. Prosper.com brings together crowds of potential borrowers and lenders from around the world and is based in the U.S. On the platform, the lending process begins when a borrower asks for a loan at a target amount and interest rate, i.e., they create a campaign. Lenders can decide to fund any fraction of the overall loan amount requested. This is done through a bidding process where lenders pledge to fund any fraction of the requested loan based on their assessment of the loan’s reward of interest vs. the borrower’s creditworthiness. In the considered data, 21.92% of projects received funding. Of the funded loans with known repayment outcomes, 65.44% were repaid.

While there are common trends in the fundraising dynamics of successful campaigns, like a steep increase in the raised capital shortly after launching, campaigns remain unique from each other

in how they arrive at the desired funds. Typically, within a week, campaigns need to raise most of the sought capital to succeed. Bids usually consist of small chunks (the median contribution is \$50, see Figure 1a) and are raised in short inter-bid times (the median is less than 10 minutes, see Figure 1b). Figure 1c depicts randomly chosen trajectories of successful as well as failed campaigns and indicates that the distribution of bid amounts and inter-bid times are the two complementary components of this dynamics.

In lending institutions like banks, loan officers may be required to have formal college training in an economics-related area and receive special training to increase the chances that they make reliably accurate loan decisions. Bank loan officers base their estimates of a borrower's creditworthiness on credit rating (i.e., the letter rating assigned to one's credit score by Prosper), debt-to-equity ratio (i.e., the ratio between the amount of money that one owes to how much they earn), collateral, and personal relationship with the borrower. The crowd of lenders makes their lending decisions with information only on credit rating and debt-to-equity ratio. Because of the size of the platform, it is unlikely that lenders know borrowers personally or are familiar with borrowers' networks except in special cases [57]. Because no collateral is pledged by borrowers as surety of repayment, lenders carry the entire risk of default associated with their portion of the loan. Finally, because loans are divided into smaller pieces spread across many lenders (i.e., loans are "syndicated"), the loan decisions are inherently collective.

On Prosper.com, borrowers' credit rating is quantified in seven typical categories AA (excellent credit, 6%), A (6.62%), B (9.44%), C (14.62%), D (17.12%), E (16.14%), and HR (high risk, 29.99%). The prevalence of high-risk borrowers suggests that individuals without access to funds from traditional financial institutions use the platform. On Prosper.com, the default rate of AA loans, which are more likely to qualify for traditional lending sources, is 5.2%². By contrast, bank lending officers, presumably equipped with formal training and more detailed background data had a default rate between 1.5% and 7.4% on diverse types of loans during the same time period [1]. This suggests that collectively lenders have a non-negligible success rate compared to trained experts and do significantly better than chance in identifying good credit risks. Previous research on an early subset of the Prosper.com data indicates that lenders can attain a greater accuracy in predicting a borrower's likelihood of defaulting on a loan than models that simply use the borrower's credit score [60]. Our work expands on this line of investigation by developing indicators that tap into the collective intelligence of lenders for improved assessment of borrower creditworthiness.

4 MEASURES

Risk categories. According to financial lending research, good lending decisions occur if creditworthy borrowers receive funds and non-creditworthy borrowers are denied funds [2]. Creditworthy borrowers are those whose risk of default relative to reward is low. High-risk borrowers have low credit ratings and ask for large loans; low-risk borrowers have high credit ratings and ask for small loans. Between these endpoints are hybrids of medium risk: borrowers with low credit ratings/small loans and those with high credit ratings/large loans. Each of the four categories has an associated

base rate of accurate predictions, which equals the fraction of successful campaigns in that group. A test of collective intelligence's accuracy is whether crowds can predict funding and default better than expected by the base rate. To construct categories of risk vs. reward (Figure 2), we thus developed four categories of loan decisions. High-risk loans involve decisions about lending big amounts to borrowers with bad credit ratings. Medium-high risk (aka Medium 1 risk) loans involve decisions about lending big amounts to borrowers with good credit. Medium-low risk (aka Medium 2 risk) loans involve decisions about lending small amounts to borrowers with bad credit. Low-risk loans involve decisions about lending small amounts to borrowers with good credit. Good credit is defined as AA and A credit rating; bad credit is defined as the HR (high risk) credit category. Big and small loan sizes are defined as being above or below the median loan size found in our data.

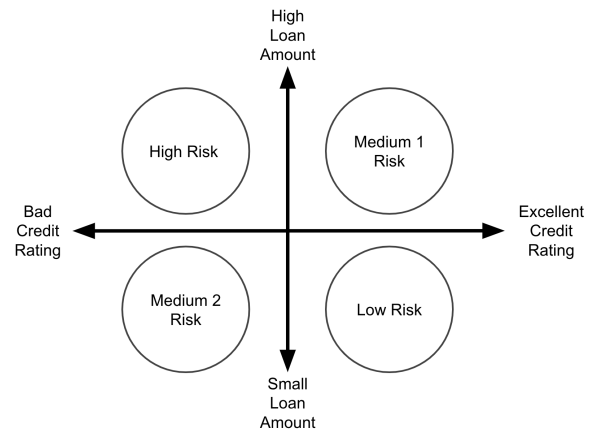


Figure 2: Illustration of risk categories in the credit rating and loan amount plane. We combine credit information from Prosper.com with loan amounts to indicate whether high/low-risk borrowers are seeking large/small amounts.

Opinion diversity. Crowdfunding lenders pledge to carry a portion of a loan based on their opinion about the borrower's creditworthiness relative to the size of the reward (repayment with interest). Investors form their opinions of the loan's risk vs. reward based on their idiosyncratic beliefs, personal and business experience, and biases [18, 21, 44, 61, 76]. This suggests that each lender's bid represents their estimate of the right risk-reward balance based on their particular information and evaluation criteria. The bidding of the crowd represents the collective diversity of the crowd. This pattern between the size of a bid and the valuation of the object or project is a basis for auction theory [101]. Bidders bid different amounts for the same object or project because they have different opinions about its worth. It may also be possible that two diverse lenders bid equal amounts, but this would only create a conservative test of our hypotheses because our diversity measure would underestimate the true diversity of the system. We computed a Gini coefficient of the sizes of lenders' bids to operationalize the opinion diversity of lenders as follows:

$$OD_X = \frac{1}{2n \sum_{k=1}^n a_X^k} \sum_{i=1}^n \sum_{j=1}^n |a_X^i - a_X^j| \quad (1)$$

²Another 10.5% of the loans are written off by the lender as a loss for unknown reasons.

where n is the number of bids and a_X^i is the amount contributed by lender i to campaign X . The coefficient ranges from 0, or complete equality, to 1, or maximum inequality. For example, a campaign receiving the amounts \$10, \$25, \$150, and \$500 has an opinion diversity of 0.58.

Aggregation speed. Building on prior research suggesting that higher aggregation speed positively correlates with successful fundraising [24, 37, 93], our second indicator quantifies how quickly lenders contribute funds by measuring the average time between bids. Thus, for each campaign X , we compute the average inter-bid time as follows:

$$AS_X = \frac{1}{n-1} \sum_{i=0}^{n-1} (t_X^{i+1} - t_X^i) \quad (2)$$

where n is the number of bids, t_X^i is the time when the i^{th} bid was made, and t_X^0 marks the time at which the campaign was launched. For ease of interpretation, we scale the average inter-bid times to the 0 to 1 interval to indicate slow to fast aggregation. For example, consider a campaign with bids that arrive 1, 15, 36, and 72 hours after the campaign's launch. The corresponding inter-bid times would be 1, 14, 21, and 36 hours, which yields an average inter-bid time of 18 hours.

Null models. We construct null models of both diversity and speed to test whether the observed data could be explained by chance, bid sizes, or inter-bid times. The tests involve taking the observed data and randomly shuffling over the observed bid sizes and inter-bid times as follows.

To compare opinion diversity with a random expectation, we compute Z-scores based on 1,000 samples of random bid amounts where bid sizes were left to chance. Each random sample consists of n uniformly distributed bid amounts taken from the actual values in the observed data that adds up to the amount funded in the campaign. The majority of observed bids are not explained by chance, even for extremely short campaigns that closed after just five bids. The bidding for 78% of the campaigns differ from what is expected by chance ($p < 0.05$ one-tailed test).

To compare aggregation speed with a random expectation, we rely on a Poisson process [6, 10] that assumes that bids occur independently at a constant rate λ and with exponentially distributed inter-bid times τ : $p(\tau) = \lambda \exp(-\lambda\tau)$. For every campaign X , we compute the theoretical expectation as the rate λX , which is equal to the total running time of the campaign divided by the number of bids it receives. We find that the difference between the expected and observed inter-bid times is statistically significant (Kolmogorov-Smirnov test: $p < 0.001$). Thus, for both collective intelligence indicators, the null models indicate that the observed data are not explained by chance.

Control variables. To evaluate the predictive power of diversity and speed, we control for a set of variables found to influence fundraising and repayment. We control for the number of bids [24], size of loan [97, 100], debt-to-income ratio [23, 42, 90], credit score [23, 42, 60], length of borrower's explanation for why they want the loan [5, 24, 65], and the state the borrower lives in. We consider the borrower's state because it can bias results

when location triggers stereotypes of persons from different regions of the U.S. or promotes in-group bias for persons living in certain states [52, 67]. For instance, 15.37% of the campaigns are from California.

5 MODELS

5.1 Selection Model

Selection bias arises in our data because funded projects are not randomly selected from all possible projects. This means information is lacking on whether an unfunded campaign would have returned the investment had it been funded. In other words, conclusions about repayment are contingent on the selection bias introduced in the stage of getting a loan. To jointly model fundraising and repayment, we thus use a two-stage Heckman probit model that considers the selection bias [53, 97]. In the first stage, we specify the selection model by estimating the probability of funding. We control for the total number of bids due to its high correlation with funding probability [24, 36] and the variables related to creditworthiness that might influence the lender's decision to fund (see Section 4, *Control variables*). In the second stage, we use the estimated Mills ratio from the first stage to account for selection bias. Additionally, we control again for the number of bids and for the two variables that directly make repayment more challenging: the amount requested and the debt-to-income ratio. In both stages, our variables of interest are the collective intelligence indicators of opinion diversity and aggregation speed.

Thus, in our two-stage selection model, the fundraising probability for a campaign X is expressed as:

$$P_X^F = I[\gamma_0 + \gamma_1 OD_X + \gamma_2 \log(AS_X) + \gamma_3 \log(b_X) + \gamma_4 a_X + \gamma_5 r_X + \gamma_6 c_X + \gamma_7 s_X + \gamma_8 d_X > 0] \quad (3)$$

whereas the repayment probabilities for a campaign X is:

$$P_X^R = \beta_0 + \beta_1 OD_X + \beta_2 \log(AS_X) + \beta_3 \log(b_X) + \beta_4 a_X + \beta_5 r_X + \delta P_X^F + \alpha \rho_X \quad (4)$$

where $\text{corr}(\gamma_0, \beta_0) = 0.25 \neq 0$ indicating that standard regression techniques applied to the regression Equation 4 would yield biased results. b denotes the total number of received bids, a is the amount requested, r is the borrower's debt-to-income ratio, c is their credit rating, s is their state, and d is the length of the proposal request. I denotes the indicator function. The parametric correction term assumes a normal hazard and equals the inverse Mill's ratio ϱ_X , which depends only on known parameters of Equation 3.

5.2 Predicting Fundraising from Small Crowds

From a theoretical and practical perspective, collective intelligence indicators are most valuable when good fundraising estimates are made early in the bidding process. To test whether such indicators can operate with small crowds, we examine whether collective intelligence indicators predicted fundraising by using a random forest classification algorithm [17] that considered the first $k = 1, 2, \dots, 10$ bids. We choose to work with random forests because they have been shown previously to perform better than other comparable algorithms in the context of crowdfunding [50]. We evaluate the algorithm's performance with 10-fold cross-validation

Table 1: Results of the two-stage Heckman model. The table shows the regression coefficients for fundraising as the selection variable and repayment as the outcome variable. The collective intelligence indicators of opinion diversity and aggregation speed are statistically significant in both cases. [†] denotes logged variables; * p<0.001**

Independent Variables	Stage 1	Stage 2
	Selection: Fundraising	Outcome: Repayment
Opinion Diversity	3.151*** (0.048)	0.340*** (0.031)
Aggregation Speed [†]	2.160*** (0.065)	0.457*** (0.039)
Number of Bids [†]	2.291*** (0.015)	0.452*** (0.014)
Amount Requested [†]	-2.813*** (0.025)	-0.586*** (0.019)
Debt-to-Income Ratio	0.027*** (0.007)	-0.014*** (0.003)
Credit Rating	✓	
Region of Borrower	✓	
Description Length	0.0002*** (0.00004)	
Constant	4.393*** (0.424)	1.467*** (0.037)
Wald χ^2		1315.49***
Log-Likelihood		-42060.43
Rho		0.454 (0.016)

using the area under the ROC curve (AUC). The AUC measures the probability that truly funded campaigns are considered by the algorithm to have better chances of successful fundraising than campaigns that failed in reality.

Beyond prediction, random forests enable us to assess variable importance. Specifically, for a chosen crowd size (i.e., a set number of bids), we first make a prediction based on the observed data. Next, we randomize the observed values of one variable. We repeat the prediction using the randomized variable and the untouched remainder of the data. The decrease in prediction accuracy quantifies the predictive power (or importance) of the randomized variable. We monitor the evolution of the importance of the variable as more and more bids are considered, which informs us about the variable's utility with changing crowd size.

6 RESULTS

Collective intelligence indicators are reliable predictors of fundraising and repayment. We begin by presenting the results of our two-stage selection model that regresses the probability of a loan being funded and the probability of a loan being repaid. As shown in Table 1, the model replicates prior work, demonstrating that the total number of received bids, the amount of capital requested, the borrower's debt-to-income ratio, as well as their credit rating, geographical and administrative region, and the length of the loan proposal request is significantly associated with who gets funded and who repays. Our model also finds that both collective intelligence indicators consistently relate to a loan being funded and repaid. We obtain similar results to the Heckman probit model using a logit regression and a random forest model to predict funding success. However, the Heckman framework has the advantage that it accounts for a crucial sample selection issue: repayment information is not available for unfunded projects. These results suggest that opinion diversity and aggregation speed are separate features unique to crowd dynamics, not substitutes for measures of a borrower's creditworthiness and default risk. Both fundraising and repayment success is more likely in the presence of high bid inequality (i.e., more opinion diversity) and low average inter-bid-time (i.e., high aggregation speed).

Collective intelligence indicators are associated with project outcome regardless of risk level. Our findings thus far show that diversity and speed predict two key outcomes related to borrowers' and lenders' crowdfunding goals but do not indicate the smartness of the crowd relative to what is expected given the background rate. Figure 3 shows the estimates of the likelihood of successful fundraising and repayment conditional on combinations of diversity and speed for four risk-reward levels grouped by low and high borrower credit rating and low and high loan amounts. The estimates are computed by adding interaction effects into the regression shown in Table 1 after dichotomizing the amount requested and credit rating as described in Section 4, *Risk categories*. In the figure generated from the Heckman selection model, the dashed horizontal red line indicates the base rates for the probability of fundraising and repayment in each of the risk categories. Three findings are noteworthy.

First, campaigns that have high diversity and high speed make better predictions about the creditworthiness of a borrower and a borrower's likelihood of default than expected by the base rate in seven out of eight risk categories. Repayment in the low-risk category is underestimated by crowd wisdom, potentially because the actual repayment is nearly 100% in this category. High-diversity and high-speed campaigns can be up to 92.9% better than predicted by the base rate.

Second, in all four fundraising risk categories (Figure 3, top panels), we see that the lift from crowd wisdom over the base is statistically significant. Even in the high-risk category, where the difference seems relatively small in real terms, the lift of the high-diversity and high-speed predictions is 37.6%. The four repayment categories indicate the most interesting and important theoretical and practical findings from our study (Figure 3, bottom panels). The lift provided by crowd wisdom is statistically and substantially greater than expected by the base rate in the two categories with the highest risk. In the high-risk category, for example, the base rate estimates a repayment in roughly 40% of the cases. In contrast, the combination of high diversity and high speed correctly predicts repayment 55.3% of the time.

Third, consistent with our hypotheses, we see that crowd wisdom consistently arises only when both diversity and speed are relatively high. Campaigns that are high on one but not both collective intelligence indicators have a success rate between 43.7% and 95.2% of the base rate, while campaigns that are low on both indicators (low diversity and low speed) have success rates up to 14.1% lower than the base rate.

Collective intelligence indicators as early warning signals. Next we show that the two collective intelligence indicators of diversity and speed work reliably also with small crowds, suggesting that the crowdfunding system's efficiency can be improved. To test the indicators' ability to predict fundraising success from a few bids relative to traditional measures of creditworthiness, we train a classifier at different bid numbers. We find that in the range of the first $k = 1, 2, \dots, 10$ bids, the random forest's performance in predicting fundraising increases from AUC=0.64 to AUC=0.76. With the increasing number of bids, the algorithm has more information about the progression of the bidding process, becoming more accurate.

Figure 4 shows the permutation importance of the collective intelligence indicators relative to credit rating, debt-to-income ratio, and description length. Only 38.6% of the campaigns receive more

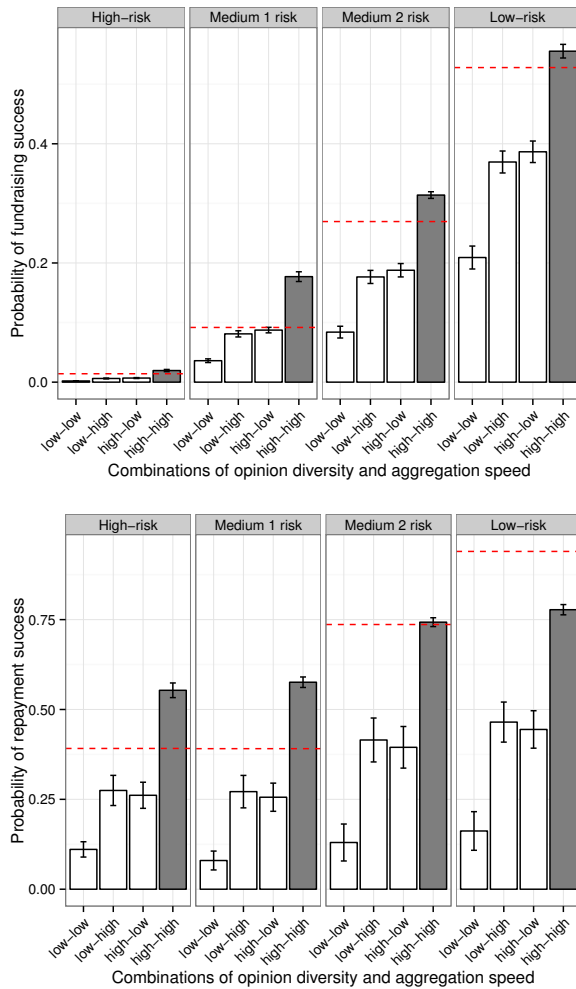


Figure 3: Probability of successful fundraising (top) and repayment (bottom) for campaigns of different risk levels, conditional on opinion diversity and aggregation speed. Dashed lines indicate base rates computed under the assumption of random campaign outcomes in each risk category; gray bars depict the high opinion diversity and high aggregation speed condition that maximizes crowd wisdom. In seven out of eight cases, these campaigns are more likely to be successful than predicted by the base rate. Campaigns with a low diversity and/or aggregation speed perform worse than the base rate.

than 10 bids and thus lie outside of this analysis. As a validity check, we observe that for the minimum value of one bid, credit rating holds the most information about project merit. However, at four bids, opinion diversity and aggregation speed have more predictive value than credit rating. The importance of credit rating decreases over the bidding, while at the same time, the predictive power of diversity and speed grows. Note that the prediction accuracy of the random forests and the importance of the collective intelligence indicators relative to credit rating can be further elevated if the number of bids is included in an algorithm that learns based on all campaigns up to a given bid number. These findings indicate that

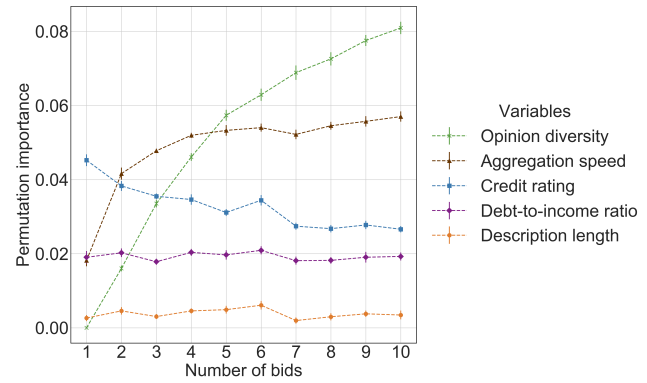


Figure 4: Collective intelligence indicators as early warning signals based on small crowds. Forecasts are made after each consecutive bid to assess prediction quality at the initial stages of campaigns. At the very beginning of the bidding process, credit rating is the most predictive variable. After only four bids, however, credit rating is outperformed by both opinion diversity and aggregation speed.

diversity and speed are powerful in predicting borrower creditworthiness in populations of non-experts making complex estimates about future unknowns.

7 DISCUSSION

Several findings about dishonesty and corruption in the banking industry [29, 98] highlight the challenges facing traditional financing models. At the same time, trends suggest that crowdfunding is growing at a rate that will possibly supplant traditional approaches from banks. Increased transparency, lack of financial intermediaries, and potential for diminished biases make crowdfunding an attractive alternative. Yet, little is known about how effective untrained crowds are at selecting creditworthy borrowers. Poor selection would misallocate funds from meritorious projects and possibly undermine the growth of crowdfunding systems.

Despite academic speculations about the ‘madness of crowds’ [72], a few studies have shown that the consequences of violating some of the conditions for collective intelligence need not be as dramatic as suspected. For instance, research on prediction markets shows that crowd performance is superior to individual decisions even when individuals rely on the observable opinions of other crowd members [7, 102]. Moreover, theoretical predictions and experimental results indicate that under certain conditions, influence among crowd members produces the most accurate judgments [4, 11, 12].

Reliance on expertise also requires extensive experience or training, which can limit the supply of experts and diversity [29, 97]. For example, up until the early 1960s, it was accepted practice by bank loan officers in the United States not to consider women eligible for a loan unless their request was cosigned by a man [22].

In this paper, we contribute to the expanding literature on the wisdom of crowds with empirical evidence of a new form of collective intelligence that the crowd is not aware of. To do so, we apply concepts from the literature on the wisdom of crowds to examine whether factors that are thought to produce collective intelligence in other contexts operate in crowdfunding using a dominant crowdfunding platform. In agreement with theoretical expectations, we find that two collective intelligence indicators, i.e., opinion diversity

and aggregation speed, quantified by the Gini coefficient of the bid amounts and the average inter-bid time, are predictive of who gets capital and who pays back. Furthermore, we find that these factors are potent at forecasting the outcome of high-risk campaigns and are indicative of successful fundraising, even in the early stages of a campaign.

Our results align with and advance work on the role of collective intelligence in crowdfunding. Others have shown recently that similar signals deduced from contribution patterns on various types of crowdfunding sites, including equity and donations-based platforms, are predictive of which projects will meet their fundraising goal [37]. This work demonstrates associations between crowd dynamics and fundraising success in different contexts. Research relying on data from Prosper.com, specifically, points to various individual factors predicting default [23, 36, 42, 60, 90], albeit without properly accounting for the selection bias, i.e., the fact that we inherently lack the knowledge of whether an unfunded campaign would have returned the investment had it been funded. Furthermore, to the best of our knowledge, our paper is the first to show that lending crowds are (1) smart in the sense that they can predict the outcome of risky loans and (2) agile, meaning that they can concur on determinations of creditworthiness after only as few as four bids. Finally, recent experimental work further demonstrates the validity of similar collective intelligence indicators across a range of conditions, such as different project types, fundraising goals, participants' interest level in the projects, their altruistic attitudes, and their susceptibility to social influence [35].

The identified collective intelligence indicators are generalizable to many crowdfunding contexts (c.f. [37] for predicting fundraising in newer datasets), simple to compute in any crowdfunding setting and distill key information from campaign dynamics. On the one hand, our indicators draw from the study of auctions, in that we focus on the amounts contributed as indicators of investors' assessment of borrower trustworthiness. Our derived concept of diversity is based on expressed opinions instead of predefined socio-demographic categories. In this regard, it is unique and one step ahead of traditional notions of diversity. On the other hand, our collective intelligence indicators exploit the temporal patterns emerging in crowdfunding by using accurate fine-grained data on campaign dynamics. Successful campaigns are characterized by lenders acting promptly, but not in unison, regarding bid amounts. The inequality of contributed capital observed in this study contrasts with hypotheses of irrational herding, which imply mimicry among lenders in risk-taking [107]. We present evidence that even though its lenders do not act independently, Prosper.com gave rise to systematic crowd wisdom with tangible performance gains. This might help us to better understand and equip the future development of the crowdfunding phenomenon, which is of interest to organizations and governments. Our findings pinpoint the key component, which is central to human computation more broadly: latent collective behaviors that enhance crowd efficiency and lay the foundations for a crowd-aware system design.

8 CONCLUSIONS

Existing research on financing, both on- and off-line, has focused on characteristics of creditworthiness, such as credit scores [2, 60], social capital [45, 49, 50, 56, 68, 97], and the presentation of the borrower [5, 18, 40]. These approaches lack a crowd-centric

perspective and thus overlook the appealing possibility that Web-based crowdfunding might have produced conditions that have enabled the emergence of superior decision-making performance at the group level. Lending crowds benefit from aggregating multiple perspectives giving a more rounded view of the risk. However, crowd members are not independent and might be restricted by the actions of those who lend first, leading to cascading behavior that may not gravitate toward rational choices. Hence the two factors that could contribute to collective intelligence in crowdfunding are (1) the diversity of the perspectives in the lending crowd and (2) the aggregation mechanism that transforms individual evaluations into collective estimations considering that the decision-makers are not acting independently.

In this paper, we developed two collective intelligence indicators tailored to the specifics of crowdfunding. First, we looked at the inequality of loans, which directly polls the diversity of a group of lenders and overcomes the weaknesses of traditional categories such as gender, education, and ethnicity. Second, to devise a suitable aggregation mechanism, we studied the evolution of a campaign and viewed the time intervals between single bids as a particular combination of opinions that reinforces the crowd's overall assessment of the merit of a project. We then showed that our two resulting indicators, opinion diversity and aggregation speed, are predictive of campaign outcomes and can admit a model that jointly explains fundraising and subsequent repayment. Moreover, these indicators work particularly well with high-risk projects and might even serve as early warning signals of funding outcomes relying on data from small crowds.

Altogether, our work has resulted in large-scale empirical evidence for collective intelligence indicators in online crowdfunding. These indicators have implications for lenders, borrowers, and platforms alike. Our findings call for more concerted digital literacy education for lenders about the positive and negative effects of crowd signals. Borrowers can improve the success of their campaigns, for example, by diversifying their outreach to multiple funder categories. Platform creators can develop higher-performing sites when harnessing collective intelligence indicators. Given that the indicators are heavily dependent on crowd activity, they could be manipulated by nefarious actors (c.f. how /r/wallstreetbets gamed the stock of GameStop [77]). Future research should therefore increase awareness of these collective intelligence indicators and investigate their robustness to potential manipulation. Finally, our findings contribute to the growing literature on collective intelligence and crowd computation, opening up new avenues for theoretical work inspired by the increasingly granular understanding of the surprising efficiency of Web-enabled wisdom of crowds systems.

ACKNOWLEDGMENTS

The work is supported by the U.S. National Science Foundation under Grant No. IIS-1755873.

REFERENCES

- [1] 2022. Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks. Online. Retrieved October 11, 2022 from <http://www.federalreserve.gov/releases/chargeoff/delallsa.htm>.
- [2] Hussein A Abdou and John Pointon. 2011. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management* 18, 2-3 (2011), 59–88.
- [3] Ajay Agrawal, Christian Catalini, and Avi Goldfarb. 2015. Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy* 24, 2 (2015), 253–274.

- [4] Abdullah Almaatouq, Alejandro Noriega-Campero, Abdulrahman Alotaibi, PM Krafft, Mehdi Moussaid, and Alex Pentland. 2020. Adaptive social networks promote the wisdom of crowds. *Proceedings of the National Academy of Sciences* 117, 21 (2020), 11379–11386.
- [5] Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- [6] Harry R. Anderson. 2003. *Fixed Broadband Wireless System Design: The Creation of Global Mobile Communications*. John Wiley & Sons, Inc., New York, NY, USA.
- [7] Kenneth J Arrow, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D Nelson, et al. 2008. The promise of prediction markets. *Science* 320, 5878 (2008), 877–878.
- [8] Niina Arvila, Heike Winschiers-Theophilus, Pietari Keskinen, Roosa Laurikainen, and Marko Nieminen. 2020. Enabling successful crowdfunding for entrepreneurs in marginalized communities. In *Proceedings of the 23rd International Conference on Academic Mindtrek*. 45–54.
- [9] Josh M Attenberg, Pagagiotis G Ipeirotis, and Foster Provost. 2011. Beat the machine: Challenging workers to find the unknown unknowns. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.
- [10] Albert-László Barabási. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435 (2005), 207–211.
- [11] Joshua Becker, Abdullah Almaatouq, and E-Á Horvát. 2020. Network Structures of Collective Intelligence: The Contingent Benefits of Group Discussion. *arXiv preprint arxiv:2009.07202* (2020). <https://doi.org/10.48550/ARXIV.2009.07202>
- [12] Joshua Becker, Devon Brackbill, and Damon Centola. 2017. Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences* 114, 26 (2017), E5070–E5076.
- [13] Joshua Becker, Ethan Porter, and Damon Centola. 2019. The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences* 116, 22 (2019), 10717–10722.
- [14] Sushil Bikhchandani, David Hirschleifer, and Ivo Welch. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades. *Journal of Political Economy* 100, 5 (1992), 992–1026.
- [15] Eric Bonabeau. 2009. Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review* 50, 2 (2009), 45.
- [16] Alessandro Bozzon, Marco Brambilla, Stefano Ceri, and Andrea Mauri. 2013. Reactive crowdsourcing. In *Proceedings of the 22nd International Conference on World Wide Web*. 153–164.
- [17] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32.
- [18] Alison Wood Brooks, Laura Huang, Sarah Wood Kearney, and Fiona E Murray. 2014. Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences* 111, 12 (2014), 4427–4431.
- [19] David V Budescu and Eva Chen. 2015. Identifying expertise to extract the wisdom of crowds. *Management Science* 61, 2 (2015), 267–280.
- [20] Gordon Burtch, Anindya Ghose, and Sunil Wattal. 2013. An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research* 24, 3 (2013), 499–519.
- [21] E. Holly Buttner and Benson Rosen. 1989. Funding new business ventures: Are decision makers biased against women entrepreneurs? *Journal of Business Venturing* 4, 4 (1989), 249–261.
- [22] Dianne G Bystrom and Barbara Burrell. 2018. *Women in the American Political System: An Encyclopedia of Women as Voters, Candidates, and Office Holders [2 volumes]*. ABC-CLIO.
- [23] Carlos Eduardo Canfield. 2018. Determinants of default in P2P lending: the Mexican case. *Independent Journal of Management & Production* 9, 1 (2018), 1–24.
- [24] Simla Ceyhan, Xiaolin Shi, and Jure Leskovec. 2011. Dynamics of Bidding in a P2P Lending Service: Effects of Herding and Predicting Loan Success. In *Proceedings of the 20th International Conference on World Wide Web (Hyderabad, India) (WWW '11)*. ACM, New York, NY, USA, 547–556.
- [25] CS Richard Chan, Annaleena Parhankangas, Arvin Sahaym, and Payat Oo. 2020. Bellwether and the herd? Unpacking the u-shaped relationship between prior funding and subsequent contributions in reward-based crowdfunding. *Journal of Business Venturing* 35, 2 (2020), 105934.
- [26] Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoung Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27, 5 (2014), 1367–1403.
- [27] Jinwook Chung and Yumin Lee. 2015. A long-term study of a crowdfunding platform: Predicting project success and fundraising amount. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. 211–220.
- [28] Robert B Cialdini. 2006. *Influence: the psychology of persuasion*, revised edition. New York: William Morrow (2006).
- [29] Alain Cohn, Ernst Fehr, and Michel André Maréchal. 2014. Business culture and dishonesty in the banking industry. *Nature* 516, 7529 (2014), 86–89.
- [30] Benjamin C Collier and Robert Hampshire. 2010. Sending mixed signals: Multi-level reputation effects in peer-to-peer lending markets. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. 197–206.
- [31] Massimo G Colombo, Chiara Franzoni, and Cristina Rossi-Lamastra. 2015. Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice* 39, 1 (2015), 75–100.
- [32] Alessandro Cordova, Johanna Dolci, and Gianfranco Gianfrate. 2015. The determinants of crowdfunding success: Evidence from technology projects. *Procedia-Social and Behavioral Sciences* 181 (2015), 115–124.
- [33] Douglas J Cumming, Gaël Leboeuf, and Armin Schwienbacher. 2015. Crowdfunding models: Keep-it-all vs. all-or-nothing. *Financial Management* (2015).
- [34] Zhi Da and Xing Huang. 2020. Harnessing the wisdom of crowds. *Management Science* 66, 5 (2020), 1847–1867.
- [35] Henry K Dambanemuya, Eunseo Choi, Darren Gergle, and E-A Horvát. 2022. Beyond Words: Experimentally Examining Hidden Crowd Signals in Online Fundraising. *arXiv preprint arxiv:2206.07210* (2022). <https://doi.org/10.48550/ARXIV.2206.07210>
- [36] Henry K Dambanemuya and Eموke-Agnes Horvat. 2019. Harnessing collective intelligence in P2P lending. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 57–64.
- [37] Henry K Dambanemuya and Eموke-Agnes Horvát. 2021. A Multi-Platform Study of Crowd Signals Associated with Successful Online Fundraising. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–19.
- [38] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. 1013–1022.
- [39] Anna Dreber, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A Nosek, and Magnus Johannesson. 2015. Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences* 112, 50 (2015), 15343–15347.
- [40] Jefferson Duarte, Stephan Siegel, and Lance Young. 2012. Trust and Credit: The Role of Appearance in Peer-to-peer Lending. *The Review of Financial Studies* 25, 8 (2012), 2455.
- [41] David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA.
- [42] Riza Emekter, Yanbin Tu, Benjamas Jirasakuldech, and Min Lu. 2015. Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics* 47, 1 (2015), 54–70.
- [43] Vincent Etter, Matthias Grossglauser, and Patrick Thiran. 2013. Launch hard or go home! Predicting the success of Kickstarter campaigns. In *Proceedings of the first ACM Conference on Online Social Networks*. 177–182.
- [44] Lisa Feeney, George H. Haines JR., and Allan L. Riding. 1999. Private investors' investment criteria: Insights from qualitative data. *Venture Capital* 1, 2 (1999), 121–145.
- [45] Seth Freedman and Ginger Zhe Jin. 2008. *Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.com*. Working Papers 08-43. NET Institute.
- [46] Francis Galton. 1907. Vox populi (the wisdom of crowds). *Nature* 75, 7 (1907), 450–451.
- [47] Elizabeth M Gerber and Julie Hui. 2013. Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 1–32.
- [48] Peter A Gloor. 2006. *Swarm creativity: Competitive advantage through collaborative innovation networks*. Oxford University Press.
- [49] Mark Granovetter. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *Amer. J. Sociology* (1985), 481–510.
- [50] Michael D. Greenberg, Bryan Pardo, Karthic Hariharan, and Elizabeth Gerber. 2013. Crowdfunding Support Tools: Predicting Success & Failure. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 1815–1820.
- [51] Douglas Guilbeault and Damon Centola. 2020. Networked collective intelligence improves dissemination of scientific information regarding smoking risks. *PLoS One* 15, 2 (2020), e0227813.
- [52] Iftekhar Hasan, Qing He, and Haitian Lu. 2017. Stereotypes in Person-to-Person Lending: Evidence from debt crowdfunding. *Available at SSRN* 3085921 (2017).
- [53] James Heckman. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. In *Annals of Economic and Social Measurement, Volume 5, number 4*. National Bureau of Economic Research, Inc, 475–492.
- [54] Ralph Hertwig. 2012. Tapping into the wisdom of the crowd—with confidence. *Science* 336, 6079 (2012), 303–304.
- [55] Lu Hong and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America* 101, 46 (2004), 16385–16389.
- [56] Eموke-Agnes Horvát, Jayaram Uparna, and Brian Uzzi. 2015. Network vs market relations: The effect of friends in crowdfunding. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. 226–233.

- [57] E. Á. Horvát, J. Uparna, and B. Uzzi. 2015. Network vs Market Relations: The Effect of Friends in Crowdfunding. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 226–233.
- [58] Julie S Hui, Michael D Greenberg, and Elizabeth M Gerber. 2014. Understanding the role of community in crowdfunding work. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 62–74.
- [59] Panagiotis G Ipeirotis and Evgeniy Gabrilovich. 2014. Quizz: targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd International Conference on World Wide Web*. 143–154.
- [60] Rajkamal Iyer, Asim Ijaz Khwaja, Erzo FP Luttmer, and Kelly Shue. 2015. Screening peers softly: Inferring the quality of small borrowers. *Management Science* 62, 6 (2015), 1554–1577.
- [61] Jenny Jordan and Klaus P Kaas. 2002. Advertising in the mutual fund business: The role of judgmental heuristics in private investors' evaluation of risk and return. *Journal of Financial Services Marketing* 7, 2 (2002), 129–140.
- [62] Rembrand Koning and Jacob Model. 2013. Experimental study of crowdfunding cascades: When nothing is better than something. Available at SSRN 2308161 (2013).
- [63] Evgeny Krivosheev, Burcu Sayin, Alessandro Bozzon, and Zoltán Szlávik. 2020. Active learning from crowd in document screening. *arXiv preprint arXiv:2012.02297* (2020).
- [64] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Eric Horvitz. 2017. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [65] Laura Larrimore, Li Jiang, Jeff Larrimore, David Markowitz, and Scott Gorski. 2011. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research* 39, 1 (2011), 19–37.
- [66] SeungHun Lee, KangHee Lee, and Hyun-chul Kim. 2018. Content-based Success Prediction of Crowdfunding Campaigns: A Deep Learning Approach. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*. Association for Computing Machinery, Jersey City, NJ, USA, 193–196. <https://doi.org/10.1145/3272973.3274053>
- [67] Karen Levy and Solon Barocas. 2017. Designing against discrimination in online markets. *Berkeley Technology Law Journal* 32, 3 (2017), 1183–1238.
- [68] Mingfeng Lin, Nagpurnanand R. Prabhala, and Siva Viswanathan. 2013. Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science* 59, 1 (2013), 17–35.
- [69] Chun-Ta Lu, Sihong Xie, Xiangnan Kong, and Philip S Yu. 2014. Inferring the impacts of social media on crowdfunding. In *Proceedings of the 7th ACM International Conference on Web Search and data mining*. 573–582.
- [70] Anna Lukkarinen, Jeffrey E Teich, Hannele Wallenius, and Jyrki Wallenius. 2016. Success drivers of online equity crowdfunding campaigns. *Decision Support Systems* 87 (2016), 26–38.
- [71] Aidan Lyon and Eric Pacuit. 2013. The wisdom of crowds: Methods of human judgement aggregation. In *Handbook of Human Computation*. Springer, 599–614.
- [72] Charles Mackay. 2012. *Extraordinary popular delusions and the madness of crowds*. Simon and Schuster.
- [73] Daniel Makina. 2019. The potential of FinTech in enabling financial inclusion. In *Extending financial inclusion in Africa*. Elsevier, 299–318.
- [74] Thomas W Malone, Robert Laubacher, and Chrysanthos Dellarocas. 2010. The collective intelligence genome. *MIT Sloan management review* 51, 3 (2010), 21.
- [75] Richard P Mann and Dirk Helbing. 2017. Optimal incentives for collective intelligence. *Proceedings of the National Academy of Sciences* 114, 20 (2017), 5077–5082.
- [76] Colin Mason and Matthew Stark. 2004. What do Investors Look for in a Business Plan? *International Small Business Journal* 22, 3 (2004), 227–248.
- [77] Norma Mendoza-Denton. 2021. "Sticking It to the Man": r/wallstreetbets, Generational Masculinity and Revenge in Narratives of our Dystopian Capitalist Age. *Anthropology Now* 13, 1 (2021), 91–99.
- [78] Tanushree Mitra and Eric Gilbert. 2014. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 49–61.
- [79] Ethan Mollick. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* 29, 1 (2014), 1–16.
- [80] Don A Moore, Samuel A Swift, Angela Minster, Barbara Mellers, Lyle Ungar, Philip Tetlock, Heather HJ Yang, and Elizabeth R Tenney. 2017. Confidence calibration in a multiyear geopolitical forecasting competition. *Management Science* 63, 11 (2017), 3552–3565.
- [81] Lev Muchnik, Sinan Aral, and Sean J. Taylor. 2013. Social Influence Bias: A Randomized Experiment. *Science* 341, 6146 (2013), 647–651.
- [82] Joaquin Navajas, Tamara Niella, Gerry Garbulsky, Bahador Bahrami, and Mariano Sigman. 2018. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour* 2, 2 (2018), 126–132.
- [83] Scott E. Page. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.
- [84] Dražen Prelec, H Sebastian Seung, and John McCoy. 2017. A solution to the single-question crowd wisdom problem. *Nature* 541, 7638 (2017), 532–535.
- [85] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Improving worker engagement through conversational microtask crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [86] Sihang Qiu, Achilleas Psyllidis, Alessandro Bozzon, and Geert-Jan Houben. 2019. Crowd-mapping urban objects from street-level imagery. In *The World Wide Web Conference*. 1521–1531.
- [87] Alexander J Quinn and Benjamin B Bederson. 2011. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1403–1412.
- [88] M. J. Salganik, P. S. Dodds, and D. J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854–856.
- [89] Ioannis Petros Samiotis, Christoph Lof, Shaad Alaka, Cynthia CS Liem, and Alessandro Bozzon. 2022. Scriptoria: A Crowd-powered Music Transcription System. In *Companion Proceedings of the Web Conference 2022*. 256–259.
- [90] Carlos Serrano-Cinca, Begona Gutierrez-Nieto, and Luz López-Palacios. 2015. Determinants of default in P2P lending. *PloS One* 10, 10 (2015), e0139427.
- [91] Shahin Sharifi Noorian, Sihang Qiu, Ujwal Gadiraju, Jie Yang, and Alessandro Bozzon. 2022. What Should You Know? A Human-In-the-Loop Approach to Unknown Unknowns Characterization in Image Recognition. In *Proceedings of the ACM Web Conference 2022*. 882–892.
- [92] Camelia Simoiu, Chiraag Sumanth, Alok Mysore, and Sharad Goel. 2019. Studying the “wisdom of crowds” at scale. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 7. 171–179.
- [93] Jacob Solomon, Wenjuan Ma, and Rick Wash. 2015. Don't wait! How timing affects coordination of crowdfunding donations. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 547–556.
- [94] Olav Sorenson, Valentina Assenova, Guan-Cheng Li, Jason Boada, and Lee Fleming. 2016. Expand innovation finance via crowdfunding. *Science* 354, 6319 (2016), 1526–1528.
- [95] James Surowiecki. 2005. *The wisdom of crowds*. Anchor.
- [96] Philip E Tetlock. 2017. Expert political judgment. In *Expert Political Judgment*. Princeton University Press.
- [97] Brian Uzzi. 1999. Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review* (1999), 481–505.
- [98] Marie Claire Villeval. 2014. Behavioural economics: Professional identity can increase dishonesty. *Nature* 516, 7529 (2014), 48–49.
- [99] Silvio Vismara. 2018. Information cascades among investors in equity crowdfunding. *Entrepreneurship Theory and Practice* 42, 3 (2018), 467–497.
- [100] Nir Vulkan, Thomas Astebro, and Manuel Fernandez Sierra. 2016. Equity crowdfunding: A new phenomena. *Journal of Business Venturing Insights* 5 (2016), 37 – 49.
- [101] Robert Wilson. 1992. Strategic analysis of auctions. *Handbook of game theory with economic applications* 1 (1992), 227–279.
- [102] Justin Wolfers and Eric Zitzewitz. 2004. Prediction markets. *Journal of Economic Perspectives* 18, 2 (2004), 107–126.
- [103] Anita Williams Woolley, Ishani Aggarwal, and Thomas W. Malone. 2015. Collective Intelligence and Group Performance. *Current Directions in Psychological Science* 24, 6 (2015), 420–424.
- [104] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (2010), 686–688.
- [105] Anbang Xu, Xiao Yang, Huaming Rao, Wai-Tat Fu, Shih-Wen Huang, and Brian P Bailey. 2014. Show me the money! An analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 591–600.
- [106] Yahoo! 2022. Crowdfunding Market Worth US\$ 28920 million by 2028. Online. Posted 3-August-2022. Retrieved 8-October-2022 from <https://www.yahoo.com/now/crowdfunding-market-worth-us-28920-112800825.html>.
- [107] Juanjuan Zhang and Peng Liu. 2012. Rational Herding in Microloan Markets. *Management Science* 58, 5 (2012), 892–912.
- [108] Qizhen Zhang, Tengyuan Ye, Meryem Essaidi, Shivani Agarwal, Vincent Liu, and Boon Thau Loo. 2017. Predicting startup crowdfunding success through longitudinal social engagement analysis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. Association for Computing Machinery, Singapore, Singapore, 1937–1946. <https://doi.org/10.1145/3132847.3132908>
- [109] Peng Zhao, Yu-Jie Zhang, and Zhi-Hua Zhou. 2021. Exploratory machine learning with unknown unknowns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 10999–11006.