Investor Retention in Equity Crowdfunding

Igor Zakhlebin
Northwestern University
izakhlebin@u.northwestern.edu

Emőke-Ágnes Horvát
Northwestern University
a-horvat@northwestern.edu

ABSTRACT

Crowdfunding platforms promise to disrupt investing as they by-pass traditional financial institutions through peer-to-peer trans-actions. To stay functional, these platforms require a supply of investors who are willing to contribute to campaigns. Yet, little is known about the retention of investors in this setting. Using four years of data from a leading equity crowdfunding platform, we empirically study the length and success of investor activity on the platform. We analyze temporal variations in these outcomes and ex-plain patterns using statistical modeling. Our models are based on information about user's past and current investment decisions, i.e., content-based and structural similarities between the campaigns they invest in. We uncover the role of past successes and diversity of investment decisions for novice vs. serial investors. Our results inform potential strategies for increasing the retention of investors and improving their decisions on crowdfunding platforms.

KEYWORDS

Equity crowdfunding; Investing; User retention; Investment success; Portfolio diversity.

1 INTRODUCTION

Online crowdfunding platforms enable their users to contribute to proposed campaigns without the involvement of an institutional mediator between them and the people requesting funds [2, 42]. Since funding is determined by a large group of people instead of a small number of decision makers, campaigns can benefit from the aggregated wisdom of crowds, which promises to recognize valuable projects that would have remained unfunded by traditional financial organizations [9]. In only a couple of years, crowdfunding expanded from funding aspirant creative works and supplying pro-social donations to enabling large citizen-funded urban projects and providing commercial interest-based unsecured loans, as well as angel investments [22]. Due to this growth and diversification, the industry has been increasingly a subject of scholarly interest, mainly in entrepreneurship and management literature [30, 39, 50, 55].

The latest addition to various types of crowdfunding is equity crowdfunding, which allows startups to attract funds from a large group of investors in return for a stake in company ownership [3]. What makes it key among other forms of crowdfunding is that it competes with venture capital by scaling up from the "family and friends" funding model to professional financial markets [42, 50] where investors are motivated by a stake in company ownership and future financial returns instead of an immediate deliverable [13]. Given the quick expansion and current maturity of equity crowd-funding [17], this format promises to truly transform decentralized markets and is thus a crucial subject of study [26, 28].

Regardless of the specific form of crowdfunding, understanding investor participation is crucial. Crowdfunding platforms function as marketplaces that match investments with campaigns. The effi-ciency of such matching determines how well the platform fulfills its purpose, to the extent that all major platforms use the amount of money successfully invested as one of their core key perfor-mance indicators. A lack of investors makes it harder for individual campaigns to receive the requested funding, which decreases the efficiency of the entire platform and, in extreme cases, can lead to the platform becoming defunct [1].

Recent developments demonstrate that crowdfunding platforms cannot rely solely on extensive growth to fund their campaigns. A recent analysis of Kickstarter shows that the platform has reached a plateau in the amount of money pledged to successful cam-paigns [10]. Considering that the growth of this platform was histor-ically fueled by intake of new crowd investors [11], this slowdown may be attributed to the saturation of the market as most of the interested users have either already joined Kickstarter or one of its competitors. Additionally, some existing investors inherently lose interest in crowdfunding over time, a phenomenon that was dubbed "crowdfunding fatigue" by the media [44, 57].

We therefore focus on the following question in the context of equity crowdfunding: *What factors contribute to the retention of investors?* Using comprehensive data about investments made on a popular equity crowdfunding platform over the span of four years, we explore how investment decisions affect investor retention on the platform. Our main contributions can be summarized as follows:

- In light of existing literature (Section 2) and based on our data (Section 3), we introduce variables to describe individual invest-ments and assess the success of crowd investors (Section 4.1–4.2). To better understand the factors associated with investor re-tention, we construct new measures that enable exploring the usefulness of portfolio-based investment rules in the context of equity crowdfunding. Our measures capture the dissimilarity of investments within a "portfolio" with respect to their categories, descriptions, and shared investor bases (Sections 4.3-4.4).
- We show that, as expected, investors who remain on the plat-form have higher success rates than the ones who leave and that over time serial investors become less aligned with the crowds'

investment choices (Section 5.1). We also document and explain a counterintuitive uptake of success rate for investors prior to leaving the platform.

- Using survival analysis, we explain the rate of attrition over time based on the proposed variables (Section 5.2). Results show that for novice investors, an exploration of different campaigns is important for retention. Serial investors who stay longer on the platform tend to invest in tandem in the same campaigns and they are more likely to be successful when they choose significantly different campaigns from the same category.

We then discuss our results in the context of broader crowdfunding and crowdwork literature (Section 6) and conclude by outlining directions for future work (Section 7).

2 RELATED WORK

Although there are few papers that focus on the question of investor retention, some closely related questions have been explored in the following bodies of literature.

Contributor retention. Since crowdfunding is enabled by online platforms and relies on contributions from many participants, it can be regarded as a special case of peer production or crowdsourcing, where contributors bring campaign proposals and personal finances instead of ideas, opinions, or effort [22]. This allows us to leverage the rich literature about contributor retention in peer production and crowdwork that covers platforms like Wikipedia [23, 24], Wikia [48], OpenStreetMap [15], Q&A sites [16, 43, 54], forums [37], newsgroups [7, 31], and social media sites [34, 53].

The shared theme of this research is the retention of newcomers, which is strongly affected by their experience during the first contributions on the platform [24, 31, 48]. Current performance, as captured by the frequency, speed, and overall number of contributions, has a positive impact on retention of novice [15, 16, 37, 43] and continuing contributors [53].

While we expect these findings to be general enough to apply to crowdfunding, this is not a given. The types of actors and contributions on such platforms are significantly different [25]. For example, campaigns typically remain fixed once proposed and sourced money are primarily considered as a lump sum. Most importantly, unlike in peer production and crowdsourcing scenarios, there is no explicit coordination in crowdfunding.

Crowdfunding research. Literature systematizing existing research and setting future agenda has focused on classifying crowdfunding platforms and their business models [25] and outlined both similarities with previous funding models, as well as important theoretical and empirical differences from them [22]. Research on user behavior has primarily focused on the project creator's experience on crowdfunding platforms. Several studies have uncovered factors associated with the success of individual campaigns in obtaining the requested funds by assessing the role of creditworthiness [30], effect of social capital [19, 27, 29, 39], and presentation of the idea [4, 18, 28]. The most studied aspects from the investor's point of view have been the motivations for participation [20, 21, 47], the choice of campaigns [6], and the temporal dynamics of contributions towards individual campaigns [12, 36, 55].

The *only* studies related to contributor retention in crowdfunding to date are based on a single education-centric donation platform,
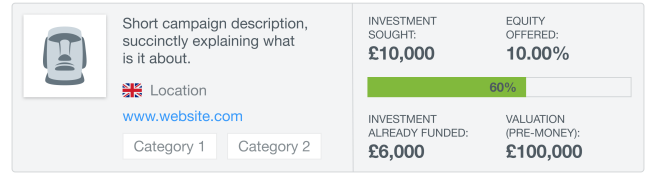

Campaign Name

Figure 1: Mock-up of platform's user interface card with information about a single campaign. The detailed description associated with the campaign is not shown here.

DonorsChoose.org [5, 51]. Our paper fills an important gap in the literature by providing a second, significantly different study of investor retention in a novel and high-stakes form of crowdfunding. We also move beyond analysis of first-time investors and develop models explaining the behavior of serial investors. This is essential since only 1% of donors on DC.org had five or more contributions, while the percentage of serial investors in our case is 11.8%.

3 DATA

We received comprehensive data from one of the leading equity crowdfunding platforms operating in the UK and the greater European area. Project creators on this platform are primarily start-ups and later-stage companies that raise capital for their subsequent rounds of funding. The platform thus favors large campaigns, the mean campaign target being £163,800. The investor side is represented both by small private and larger institutional investors. Our data includes information about all crowdfunding campaigns and all individual investments made by platform users between July 2012 and January 2016. We only consider users who have made at least one investment on the platform, and we will heretofore refer to them as "investors." The data comprises 16,907 investors who have collectively made 59,370 investments into 727 campaigns during the considered time period.

On platform's web interface, campaigns are summarised in the form of cards, displaying their requested amount, offered share in equity, current status, and other relevant information (see Fig. 1). Additionally, each campaign features a comprehensive description typically consisting of a few paragraphs of text and is tagged by its creators with one or more predefined categories, such as "Energy," "Consumer products," or "Media and Entertainment." The platform operates by an "all or nothing" principle, meaning that a campaign succeeds only if it reaches its target amount by the campaign expiration date. Otherwise, it is cancelled and all collected money are returned to investors. Campaigns in our sample have an average duration of 83 days, and only 33.6% of them are successful. Finally, each investment is associated with a time stamp and an amount in £ or €, which we convert for consistency into £ according to the exchange rate at the time of investment.

4 MEASURES AND MODELS

To describe investor participation on the platform, we use the pipeline shown in Fig. 2. Investors are presented with a variety of ongoing campaigns to choose from. Based on campaign details and previously made investments, they decide which campaign
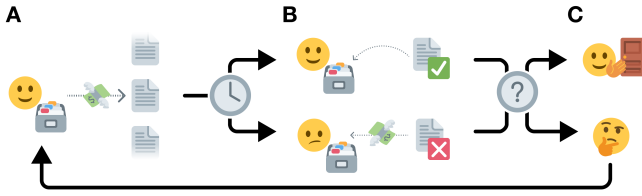
Figure 2: Pipeline of participation in equity crowdfunding. (A) The investor chooses a campaign to invest in. (B) After waiting, they learn whether the funds have been accepted and, given this outcome, (C) they either decide to leave the platform or adjust their strategy and invest again.

they will invest in and how much money they will invest. Upon each campaign's expiration, investors learn whether the requested amount was reached and, therefore, whether their investments have been accepted. Then investors decide whether to repeat this process and make an investment in another campaign or to leave the platform. Our study particularly focuses on this last step: the choice of investors to stay on the platform or to leave.

In order to make further analyses feasible, we add a simplifying assumption to our model. Instead of allowing multiple pending investments, we require investors to decide about new contributions only after they find out whether their previous investment was successful or not. This assumption is supported by research that found previous activity to impact user retention on various platforms [24, 54]. To formalize the intuition that the behavior of individuals changes in response to the outcome of their actions, we thus assume that decisions about ongoing participation on the platform are made successively after learning about the success of the prior investment [8].

In what follows, we examine a number of measures previously shown in the context of peer production and crowd donations to affect the retention of contributors. We additionally introduce measures of campaign dissimilarity to evaluate *(1)* how diverse the portfolio is thus far and *(2)* how novel an investment is given the preceding investments. To estimate the effects of the selected variables on the retention of investors, we fit Cox's proportional hazard models [14]. The code for these computations is provided on GitHub[1].

4.1 Quantifying investor success

Just like other platforms involving crowd contributors, crowdfunding platforms benefit from users who participate for long periods of time and make several investments. To capture the longevity of investors through their ongoing interaction with the platform, we look at the *number of investments* they have made throughout the period of study. This variable allows us to track investor attrition from the platform and parallels the number of contributions in other studies [53].

Investors are interested in supporting campaigns that hit their target to avoid locking up funds until a campaign's expiration date. Such investments incur a cost of missed opportunity. To quantify how successful each user is at investing, we measure their *success*

rate, i.e., the fraction of times the user has invested in a successful campaign. It has been shown that crowd investors who invest in successful campaigns are more likely to return to the platform [5]. We are interested in testing this effect based on our data and learning how success rate influences retention of investors in the studied equity crowdfunding platform.

4.2 Characterizing individual investments

Our aim is to capture different aspects of investment behavior that may correlate with sustained contributions on the platform. Relying on existing literature, we use the following variables to describe each individual investment:

Amount. Investment amount in £. Larger investments may indicate investors' higher involvement with the platform and could influence retention similar to effort exerted by contributors in peer production [15].

Time gap. Number of days passed since the previous investment in a given campaign. For the first investment, we define time gap as the time passed since the start of the campaign. Similarly to observations of other online resources [40, 43], we expect that users who take longer to invest again are less likely to return.

Campaign stage. Fraction of the requested capital amassed by a campaign before the user invests in it. Previous studies have used similar measures to separate investors into "early movers" and followers and found important distinctions between them in terms of their rates of return [5] and their contribution to a campaign's success [36, 38].

4.3 Evaluating the dissimilarity of campaigns

To understand how individual campaigns chosen by an investor relate to each other, we quantify their dissimilarity based on salient campaign characteristics as follows:

Category-based dissimilarity. Proposers on our platform tag each campaign with one or more preset category that can be used to define coarse-grained differences between campaigns. Based on these categories, each campaign i can be described by a bit vector $x^{(i)}$ whose elements are equal to 1 if the campaign belongs to the corresponding category and 0 otherwise. We use the *Jaccard distance* to compare two campaign vectors $x^{(i)}$ and $x^{(j)}$:

$$\delta^{\text{categ}}(x^{(i)}, x^{(j)}) = 1 - \frac{x^{(i)} \cdot x^{(j)}}{\sum x^{(i)} + \sum x^{(j)} - x^{(i)} \cdot x^{(j)}}.$$

Here, $x^{(i)} \cdot x^{(j)}$ denotes the element-wise (dot) product of two campaign vectors and $\sum x^{(i)}$ denotes the number of category labels associated with campaign i.

Lexical dissimilarity. It has been shown that linguistic content is important in contexts when one needs to convince others to contribute to a campaign [4, 28]. We use the full-text description of campaigns to make finer-grained distinctions between them based on specific words used. To construct vector representations of these campaign descriptions, we apply a number of common transformations to them. First, we split them into tokens and filter out tokens that represent punctuation and common English words[2]. Then, we stem the remaining tokens, enumerate them, and build vectors with

the tokens' corresponding *TF-IDF* coefficients [32]. We start with a set of campaign descriptions D and a set of tokens T. Denoting the count of appearances of token t in campaign description d with $f_{t,d}$, we derive a vector $x^{(d)}$ to describe that campaign as follows:

$$x_t^{(d)} = TF(t,d) \cdot IDF(t,D) = \frac{f_{t,d}}{|d|} \cdot \log \frac{|D|}{d \in D:\, t \in d}.$$

Here, $|d|$ denotes the number of tokens in campaign description d and $|D|$ is the total number of campaigns. Then, to compute the dissimilarity of these vectors, we use the standard *cosine distance* as follows:

$$\delta^{\text{lex}}(x^{(i)}, x^{(j)}) = 1 - \cos(x^{(i)}, x^{(j)}) = 1 - \frac{x^{(i)} \cdot x^{(j)}}{||x^{(i)}||\, ||x^{(j)}||}.$$

Again, $x^{(i)} \cdot x^{(j)}$ denotes the element-wise (dot) product of two vectors and $||x^{(i)}||$ is the Euclidean norm of a vector, i.e., the square root of its elements' sum of squares.

Structural dissimilarity. In equity crowdfunding, it has been shown that initial contributions by investors with public profiles can increase subsequent contributions from both early and late-stage investors [49]. Emerging groups of investors who systematically invest in the same campaigns, i.e., who co-invest, could thus be correlated both with campaign success and investor participation. To quantify this association, we measure the (dis)similarity of campaigns based on investors who participate in them. Specifically, we use the principle of structural equivalence from social network analysis literature [52]. This principle posits that two nodes connected to the same others are structurally equivalent to one another. A corollary to this principle is that the amount of common neighbors in a network reveals how similar two nodes are. Analogously, campaigns that have a large overlap in investors are considered to be more similar to each other than the ones having a small or no overlap at all. To quantify this similarity, we construct an investor-campaign adjacency matrix, where each binary entry indicates whether an investor has contributed to a certain campaign before a given time point. To compare campaigns, we measure the cosine distance between their corresponding adjacency vectors, as defined above.

Viewed jointly, category-based and lexical dissimilarity allow us to evaluate the distance between campaigns based on their topic on a coarse level (*Do they belong to the same categories?*) and on a fine-grained level (*Was the wording in their descriptions similar?*). Structural dissimilarity evaluates campaigns' shared investor base, indicating whether collective behavior and personal influence might be contributing to the decision of staying on the platform or leaving.

4.4 Defining portfolio diversity and investment novelty

Having different notions of campaign dissimilarity, we adapt portfolio-based investment approaches to the crowdfunding scenario. Similarly to traditional investing, it can be assumed that new investments are decided on with existing ones in mind, which together form a *portfolio*. Modern portfolio theory postulates that owning different kinds of financial assets is less risky than owning one type alone [41]. In particular, having a portfolio of assets with uncorrelated returns allows investors to mitigate the risk posed by the potential failure or underperformance of individual assets. Since investors in equity crowdfunding typically don't have historical

data on the previous performance of companies, they try to select a diverse set of campaigns by other means. They can assess dissimilarity based on campaign descriptions and signals from other investors.

We estimate the diversity of a portfolio using the *average pairwise distance* from Ziegler et al. [56]. This measure of diversity can be considered a particular case of the Rao diversity coefficient when elements (here campaigns) are equally probable [45]. Given a set of descriptions of n campaigns $\mathbf{X} = \{x^{(1)}, \dots, x^{(n)}\}$ that a user has already invested in and a distance or dissimilarity measure δ defined on the elements of $\mathbf{X}$, we have:

$$APD(\mathbf{X}) = \frac{1}{n(n-1)/2} \sum_{i=1}^{n} \sum_{j=i+i}^{n} \delta(x^{(i)}, x^{(j)}).$$

There is a problem with using *APD* as a measure of diversity. Since we utilize it for predicting future events, we re-compute *APD* with every new investment based on information about investments made up to that time point. For the initial investments, thus, the portfolio is comprised of few campaigns, meaning that *APD* is heavily influenced by potential outliers. To mitigate their influence, we add m "fake" distance terms to each computation that are equal to the average pairwise distance $\overline{\delta}$ across all campaigns that were co-invested into in the entire dataset. This way, we incorporate our prior belief that average distances should be close to the mean distance across the population [35], which yields the *Bayesian average pairwise distance* defined as:

$$BAPD(\mathbf{X}) = \frac{1}{m+n(n-1)/2} (m\overline{\delta} + \sum_{i=1}^{n} \sum_{j=i+i}^{n} \delta(x^{(i)}, x^{(j)})).$$

To quantify investment novelty, we can similarly measure the difference between the new campaign x' that the user invests in and all campaigns he or she has previously invested in. For that, we average the pairwise distances between x' and all the prior campaigns $\mathbf{X}$:

$$BAD(\mathbf{X}, x') = \frac{1}{m+n} (m\overline{\delta} + \sum_{i=1}^{n} \delta(x^{(i)}, x')).$$

We find that the median number of n across all computations of diversity and novelty is 4. Based on that, we use $m = 4$ when computing *BAD* and $m = {}^{4(4-1)}/{}_2 = 6$ when computing *BAPD*. Given a portfolio of $n = 4$ campaigns, the weights of actual dissimilarities in the average would sum up to $^1/_2$ and grow for larger values of n.

Having these measures as a general framework, we compute diversity and novelty using the vector representations of campaigns and the introduced dissimilarity measures (see Section 4.3). This gives us three definitions of portfolio diversity ($BAPD^{\text{categ}}$, $BAPD^{\text{lex}}$, and $BAPD^{\text{struct}}$) as well as three definitions of investment novelty (BAD^{categ}, BAD^{lex}, and BAD^{struct}). These measures capture the diversity of the portfolio and the novelty of the current investment given previous investments along three different dimensions. Figures 3 and 4 show the correlation and variance relative to the mean for the category-based, lexical, and structural versions of diversity and novelty, respectively. Accordingly, the correlation between diversity and novelty is typically weak despite the same underlying formulations of dissimilarity. The variance relative to the mean is limited for both measures. The very low variance in *BAPD* in case of the first few investments makes this measure less suited to study novice investors. For this group, *BAD* has an appropriate sensitivity due to its focus on single new investments.

When studying investor retention, we use *BAPD* to measure *portfolio diversity* and *BAD* to quantify a new campaign's dissimilarity to the existing portfolio, i.e., *investment novelty*. We hypothesize
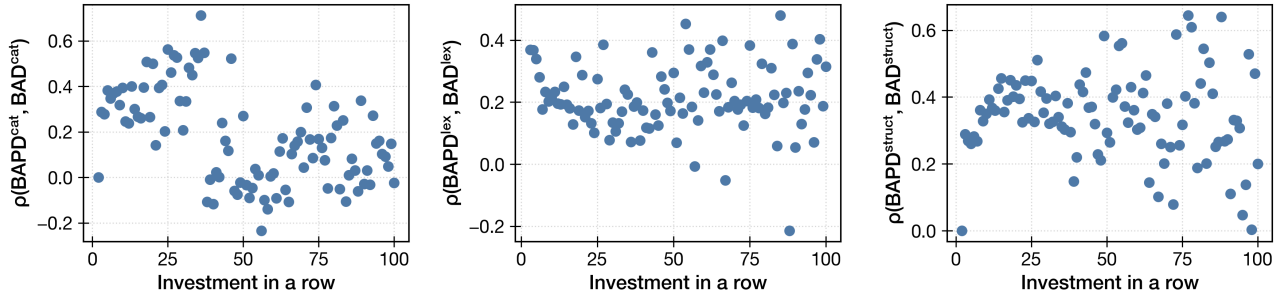
Figure 3: Pearson correlations between *BAPD* and *BAD* measures, by investment number.
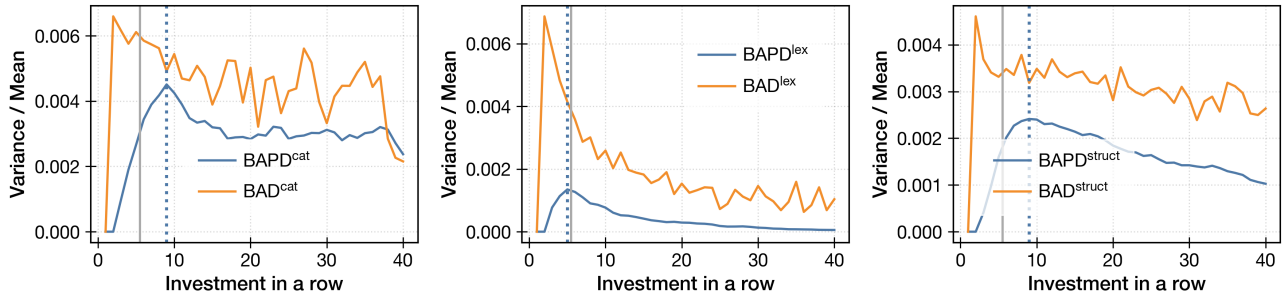

Figure 4: Variance of different *BAPD* and *BAD* measures relative to their means, by investment number. Grey line separates the first five investments from the rest, and blue dotted line shows the maximum of *BAPD*.

that investors with higher portfolio diversity are more likely to make successful investments and to stay on the platform. The expected effect of investment novelty on retention is less clear. On the one hand, it could increase the probability of leaving the platform, because if the chosen campaign is unlike the previous ones, the investor is probably less experienced with the new investment and is making a rather risky decision. On the other hand, it could help with retention, because it shows an interest that makes the investor overcome the cost of including a dissimilar campaign to their portfolio [46].

4.5 Analyses

We consider that users have left the platform if they have not made an investment at least six months prior to the end of our data collection[3]. If a user has made an investment in that period, we presume that they are still active. In statistical terms, the information about the moment when they leave the platform is *right-censored*. We exclude from our analysis users who have joined the platform in the last six months. According to these definitions, 61.5% of users have left the platform; for 9.3% the outcome is right-censored; and 29.3% are excluded as late-joiners. We estimate the user survival function using the Kaplan-Meyer method [33] and find that the median life expectancy of a user is 66.5 days or two investments.

Data preprocessing. Most platform users invest once in a campaign and occasionally they reinvest in the same campaign. Some investors, however, repeatedly target the same set of campaigns. Since this behavior can point to attempts by campaign creators to generate visibility of traction or can even represent automated investment activity, we exclude users for whom the fraction of re-investments was above 50%. To additionally control for re-investments into the same campaigns, we add a dummy variable *repeated investment* into all models: Its value is 1 if the investor has previously contributed to that campaign and 0 otherwise.

Survival models. We build a number of survival models to estimate the hazard rate of user attrition based on the variables associated with individual's investment events. Since covariates in these models change with time, we use Cox's proportional hazard model with time-varying covariates [14]. As an independent variable, we use the occurrence of an event (leaving the platform) by the end of our time period: 1 if the user has left and 0 if the information about them is right-censored and their outcome is unknown.

We fit two sets of models. The first models are for investments 1 to 5 (novice investors), and the second models are for investment 6 and higher (serial investors). We do this separation for two reasons. If we look at the hazard rate, the probability that an investor would leave the platform after each individual investment (Fig. 5), we see that there is a change around investment 5. Most investors leave the platform after the first few investments (high hazard rate), while investors who placed more than a few investments have a much lower probability of leaving (low hazard rate). This indicates that
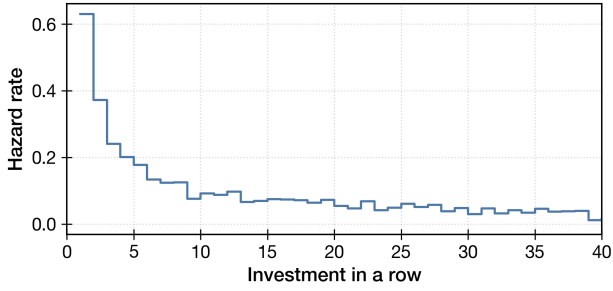
---

[3]We replicated our results with different time windows, including 3 months and 1 year, which yielded similar results.

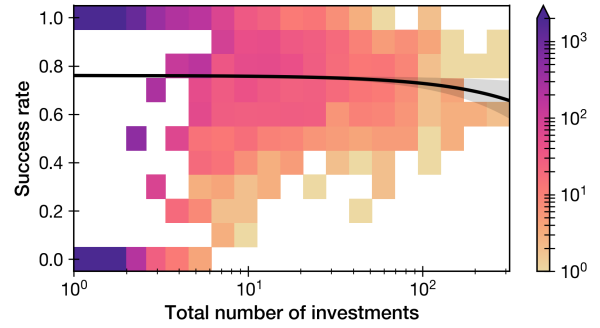Figure 5: Hazard rates for subsequent investments.

Figure 6: Distribution of success rates with respect to the total number of investments. Color denotes the number of investors in each bin and regression line shows the best linear fit to the data.

different factors can lead to novices and serial investors leaving the platform. Fitting separate sets of models also allows us to see the changes in effect sizes and significance between these two cases. Our diversity variable naturally supports this separation as well given that the measure reaches its maximum relative variance close to the sixth investment (c.f. Figure 4). We construct multiple models for both novice and serial investors including individual campaign variables along diversity and/or novelty measures.

5 RESULTS

5.1 Success rate and retention

Overall, success rates tend to remain stable over time with a slight downward trend (Fig. 6). Success rates for investors who left the platform after many investments tend to be lower than for the ones who made few investments. Figure 7, left shows how success rate changes over time with each subsequent investment for the entire population of platform users. It similarly has a downward trend showing that over time users tend to make more investments into less successful campaigns. When we split investors with respect to attrition status (Fig. 7, center), we see that investors who leave the platform at later stages are less successful than the ones who remained. This observation hints at the lower success rate as possible reason for user attrition. When we re-scale the plot with investors' final number of investments (Fig. 7, right), we find that the success rate of the last investment for leaving investors is higher than the success rate of their first investment, something we do not observe for remaining users (not shown). This observation points to a tendency of users to leave after a successful investment.

5.2 Survival analyses

The results of fitting Cox's survival models to our data are presented in Tables 1 and 2. In Models 1–4, which describe the first 5 investments (Table 1), almost all single-investment variables are significant. The investment amount and repeated investments are positively associated with attrition. These are most likely users who are trying to leave the platform. They might have been drawn to the platform by a specific campaign and do not explore further opportunities. Additionally, the more time that passes since the previous investment, the more time it takes the investor to return to the platform. Finally, if the current investment is successful and is placed late in the campaign, the investor is more likely to leave.

Table 1: Coefficients of Cox's proportional hazards model for novice investors, i.e., based on the first 5 investments.

	Model 1	Model 2	Model 3	Model 4
amount	0.03**	0.02*	0.03**	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)
time gap	0.17***	0.16***	0.19***	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)
campaign stage	−0.14***	−0.13***	−0.15***	−0.14***
	(0.01)	(0.01)	(0.01)	(0.01)
repeated inv.	0.10***	0.05**	0.08***	0.04*
	(0.01)	(0.02)	(0.02)	(0.02)
inv. success	0.09***	0.09***	0.09***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
categ. novelty	−0.11***			0.04
	(0.02)			(0.02)
lex. novelty		−0.18***		−0.21***
		(0.02)		(0.03)
struct. novelty			−0.12***	−0.01
			(0.02)	(0.03)
observations	24, 451	24, 451	24, 451	24, 451
investors	12, 742	12, 742	12, 742	12, 742
lik. ratio	558.84	637.79	534.45	640.5

Note: $* \, p < 0.05$; $** \, p < 0.01$; $*** \, p < 0.001$.

When included in the model individually, all novelty variables are negatively associated with attrition and are highly significant (Models 1–3). This indicates that novice investors profit from choosing campaigns that are dissimilar to the ones they selected before. Our finding underscores the importance of exploration in the retention of novice investors. When portfolio distance measures are combined (Model 4), only lexical distance remains significant.

For serial investors (Models 5–8), we obtain different results. The effect of time gap on attrition remains positive and highly significant, and repeated investments become occasionally significant with the same sign. Investment amount and campaign stage,
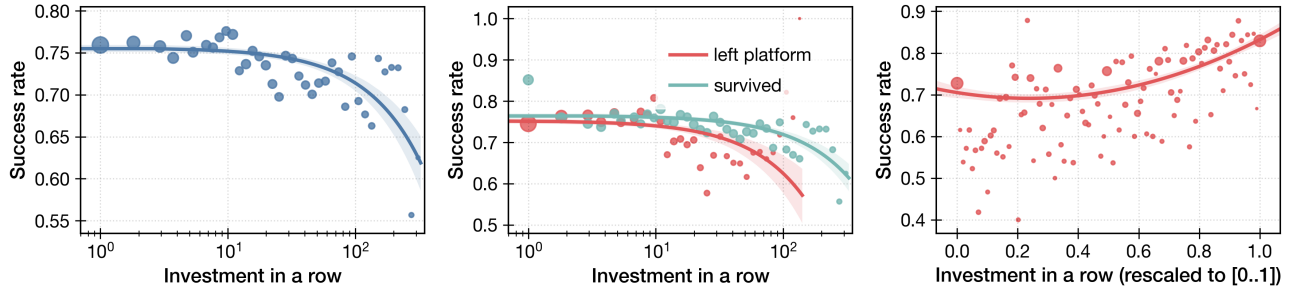
Figure 7: Success rates of subsequent investments overall (left), separated by whether investor has stayed on the platform (center), and for investors who have left (right). Solid lines represent best linear (left and center) or quadratic (right) fit to the data and shaded areas show the 95% confidence intervals. Areas of bubbles reflect the number of aggregated data points.

however, change sign: the less money invested and the later in the campaign, the more likely that this investment would be the last one. This indicates that, in order to stay, serial investors need to invest different amounts, at different campaign stages in comparison with novice investors.

With respect to diversity and novelty measures, category novelty is not significant, but category diversity is significant and positive. Users who have invested in campaigns across different categories are more likely to leave than users who have only participated in one or a few. Lexical novelty and diversity are negative and significant only in the combined Model 8: the more dissimilar campaigns users add to their portfolios, the more likely they are to continue investing. As for structural variables, they are both positive and significant. This indicates that "trail blazers" are at a higher risk of attrition. Naturally, when there is less confirmation and support through other investors who would have also recognized merit, campaigns fail to hit their target and that penalizes investors.

6 DISCUSSION

Our results provide novel explorations into the overlooked parallels between crowdfunding and peer production, suggesting new ways to improve retention of investors on equity crowdfunding platforms.

Retention of contributors. As in peer production, our results demonstrate the existence of a feedback loop in the number of contributions: with more investments, the probability to remain on the platform steadily increases [53]. In other words, users contributing to campaigns less frequently have a significantly higher rate of attrition, mimicking trends found on question answering sites [16]. Similar to the conclusions of Althoff & Leskovec [5], investors with higher success rates are more likely to remain on the platform than investors with lower success rates. However, in our data, investors become less aligned with mainstream investment decisions. We also find that investors continue investing until they make a contribution to at least one funded campaign. In line with previous research about rational herding, preexisting relations between users and campaigns are important for user retention [55]. Based on the introduced measures, we also find that both lexical and structural novelty and diversity are important factors for the retention of

Table 2: Coefficients of Cox's proportional hazards models for serial investors, i.e., based on investment 6 and higher.

	Model 5	Model 6	Model 7	Model 8
amount	−0.10∗	−0.12∗∗	−0.09∗	−0.11∗∗
	(0.04)	(0.04)	(0.04)	(0.04)
time gap	0.57∗∗∗	0.55∗∗∗	0.51∗∗∗	0.48∗∗∗
	(0.04)	(0.04)	(0.04)	(0.04)
campaign stage	0.06	0.07	0.11∗	0.16∗∗
	(0.05)	(0.05)	(0.06)	(0.06)
repeat inv.	−0.02	−0.06	0.24∗∗∗	0.15∗
	(0.06)	(0.06)	(0.06)	(0.07)
inv. success	0.28∗∗∗	0.28∗∗∗	0.32∗∗∗	0.29∗∗∗
	(0.06)	(0.06)	(0.06)	(0.06)
categ. novelty	−0.07			−0.00
	(0.05)			(0.06)
categ. diversity	0.10∗			0.17∗∗
	(0.04)			(0.05)
lex. novelty		−0.07		−0.13∗∗
		(0.04)		(0.05)
lex. diversity		−0.03		−0.10∗∗
		(0.03)		(0.03)
struct. novelty			0.37∗∗∗	0.43∗∗∗
			(0.07)	(0.07)
struct. diversity			0.13∗	0.11∗
			(0.05)	(0.05)
observations	23, 648	23, 648	23, 648	23, 648
investors	1, 476	1, 476	1, 476	1, 476
lik. ratio	254.53	255.91	299.48	332.77

Note: ∗ $p < 0.05$; ∗∗ $p < 0.01$; ∗∗∗ $p < 0.001$.

serial investors. Additionally, we find that all forms of investment novelty are beneficial for retention of novice investors.

Implications for equity crowdfunding. We find that there are two distinctive stages in investor participation. Novice investors should be exposed to a wider variety of campaigns, while serial investors should choose their target categories, settle in them, and monitor the activity of similar investors who might have already tried to

add the new combination of campaigns to their portfolios. Our findings could be leveraged in a recommender system, similarly to the proposition of An et al. [6].

Future work. Our research reiterates the role of investors in the crowdfunding process and motivates further research into contributor retention. Specifically, our observational study could be supplemented by controlled experiments that could establish causal effects between the considered variables. Additionally, similarly comprehensive data from several other crowdfunding platforms could test differences in attrition rates as function of platform design and regulations.

7 CONCLUSION

Our research contributes to the existing body of knowledge about crowdfunding by analyzing the retention of investors in an equity crowdfunding platform. In particular, we used empirical data from a market-leading platform to build models that explain investor retention based on information about their individual investments and dissimilarities between the chosen campaigns. Our results demonstrate an exploration-exploitation trade-off in investor retention: in the initial stages, investors are more likely to stay on the platform if they are exposed to a diverse set of campaigns. For serial investors, stability in terms of category and having choices aligned with other serial investors matters. These findings provide new knowledge about how investors can improve the growth prospects of crowdfunding platforms and thereby contribute to the democratization of investment opportunities online.

ACKNOWLEDGMENTS

REFERENCES

[1] [n. d.]. Crowdfunding Platforms (Defunct) - Wikipedia. https://en.wikipedia.org/wiki/Category:Crowdfunding_platforms_(defunct).
[2] Ajay Agrawal, Christian Catalini, and Avi Goldfarb. 2014. Some Simple Economics of Crowdfunding. *Innovation Policy and the Economy* 14, 1 (2014), 63–97. https://doi.org/10.1086/674021
[3] Gerrit K.C. Ahlers, Douglas Cumming, Christina Günther, and Denis Schweizer. 2015. Signaling in Equity Crowdfunding. *Entrepreneurship Theory and Practice* 39, 4 (2015), 955–980. https://doi.org/10.1111/etap.12157
[4] 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 *Proceedings of ICWSM'14*. 12–21.
[5] Tim Althoff and Jure Leskovec. 2015. Donor Retention in Online Crowdfunding Communities: A Case Study of DonorsChoose.Org. In *Proceedings of WWW'15*. 34–44. https://doi.org/10.1145/2736277.2741120
[6] Jisun An, Daniele Quercia, and Jon Crowcroft. 2014. Recommending Investors for Crowdfunding Projects. In *Proceedings of WWW'14*. 261–270. https://doi.org/10.1145/2566486.2568005
[7] Jaime Arguello, Brian S. Butler, Lisa Joyce, Robert Kraut, Kimberly S. Ling, and Xiaoqing Wang. 2006. Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities. In *Proceedings of CHI'06*. 959–968. https://doi.org/10.1145/1124772.1124916
[8] Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. 2004. Using Social Psychology to Motivate Contributions to Online Communities. In *Proceedings of CSCW'04*. 212–221. https://doi.org/10.1145/1031607.1031642
[9] Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher. 2014. Crowdfunding: Tapping the Right Crowd. *Journal of business venturing* 29, 5 (2014), 585–609. https://doi.org/10.1016/j.jbusvent.2013.07.003
[10] Thomas Bidaux. 2018. Kickstarter in 2017 - Year in Review. http://icopartners.com/2018/01/kickstarter-2017-year-review/.
[11] Salvador Briggman. 2015. Is Kickstarter's Growth Slowing Down? http://www.crowdcrux.com/kickstarters-growth-slowing/.
[12] 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 WWW'11*. 547–556. https://doi.org/10.1145/1963405.1963483
[13] Magdalena Cholakova and Bart Clarysse. 2015. Does the Possibility to Make Equity Investments in Crowdfunding Projects Crowd out Reward-Based Investments? *Entrepreneurship Theory and Practice* 39, 1 (2015), 145–172. https://doi.org/10.1111/etap.12139
[14] D. R. Cox. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 2 (1972), 187–220.
[15] Martin Dittus, Giovanni Quattrone, and Licia Capra. 2016. Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale. In *Proceedings of CSCW'16*. 108–118. https://doi.org/10.1145/2818048.2819939
[16] Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor. 2012. Churn Prediction in New Users of Yahoo! Answers. In *Proceedings of WWW'12 Companion*. 829. https://doi.org/10.1145/2187980.2188207
[17] Will Drover, Matthew S. Wood, and Andrew Zacharakis. 2017. Attributes of Angel and Crowdfunded Investments as Determinants of VC Screening Decisions. *Entrepreneurship Theory and Practice* 41, 3 (2017), 323–347. https://doi.org/10.1111/etap.12207
[18] Jefferson Duarte, Stephan Siegel, and Lance Young. 2012. Trust and Credit: The Role of Appearance in Peer-to-Peer Lending. *Review of Financial Studies* 25, 8 (2012), 2455–2484. https://doi.org/10.1093/rfs/hhs071
[19] Seth Freedman and Ginger Zhe Jin. 2008. Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.Com. (2008). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1936057
[20] Elizabeth M. Gerber and Julie Hui. 2013. Crowdfunding: Motivations and Deterrents for Participation. *ACM Transactions on Computer-Human Interaction* 20, 6 (2013), 1–32. https://doi.org/10.1145/2530540
[21] Elizabeth M Gerber, Julie S Hui, and Pei-Yi Kuo. 2012. Crowdfunding: Why People Are Motivated to Post and Fund Projects on Crowdfunding Platforms. *Proceedings of CSCW'12* (2012), 10.
[22] Rob Gleasure and Joseph Feller. 2016. Emerging Technologies and the Democratisation of Financial Services: A Metatriangulation of Crowdfunding Research. *Information and Organization* 26, 4 (2016), 101–115. https://doi.org/10.1016/j.infoandorg.2016.09.001
[23] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity is Causing its Decline. *American Behavioral Scientist* 57, 5 (2013), 664–688. https://doi.org/10.1177/0002764212469365
[24] Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. In *Proceedings of WikiSym'11*. 163–172. https://doi.org/10.1145/2038558.2038585
[25] Joachim Hemer. 2011. *A Snapshot on Crowdfunding*. Technical Report R2/2011. Fraunhofer ISI, Karlsruhe.
[26] Lars Hornuf and Matthias Schmitt. 2016. Success and Failure in Equity Crowdfunding. *CESifo DICE Report* 14, 2 (2016), 16–22.
[27] Emőke-Ágnes Horvát, Jayaram Uparna, and Brian Uzzi. 2015. Network vs Market Relations: The Effect of Friends in Crowdfunding. In *Proceedings of ASONAM'15*. 226–233. https://doi.org/10.1145/2808797.2808904
[28] Emőke-Ágnes Horvát, Johannes Wachs, Rong Wang, and Anikó Hannák. 2018. The Role of Novelty in Securing Investors for Equity Crowdfunding Campaigns. In *Proceedings of HCOMP'18*. 50–59.
[29] Julie S. Hui, Michael D. Greenberg, and Elizabeth M. Gerber. 2014. Understanding the Role of Community in Crowdfunding Work. In *Proceedings of CSCW'14*. 62–74. https://doi.org/10.1145/2531602.2531715
[30] Rajkamal Iyer, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue. 2009. Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? (2009). https://doi.org/10.2139/ssrn.1570115
[31] Elisabeth Joyce and Robert E. Kraut. 2006. Predicting Continued Participation in Newsgroups. *Journal of Computer-Mediated Communication* 11, 3 (2006), 723–747. https://doi.org/10.1111/j.1083-6101.2006.00033.x
[32] Dan Jurafsky and James H Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall, Pearson Education International.
[33] E. L. Kaplan and Paul Meier. 1958. Nonparametric Estimation from Incomplete Observations. *J. Amer. Statist. Assoc.* 53, 282 (1958), 457–481. https://doi.org/10.1080/01621459.1958.10501452
[34] Marcel Karnstedt, Matthew Rowe, Jeffrey Chan, Harith Alani, and Conor Hayes. 2011. The Effect of User Features on Churn in Social Networks. In *Proceedings of WebSci'11*. 1–8. https://doi.org/10.1145/2527031.2527051
[35] Daphne Koller and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques.* MIT press.

[36] Venkat Kuppuswamy and Barry L. Bayus. 2013. Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. *SSRN Electronic Journal* (2013). https://doi.org/10.2139/ssrn.2234765

[37] Cliff Lampe and Erik Johnston. 2005. Follow the (Slash) Dot: Effects of Feedback on New Members in an Online Community. In *Proceedings of GROUP'05*. ACM Press, 11. https://doi.org/10.1145/1099203.1099206

[38] Yan Li, Vineeth Rakesh, and Chandan K. Reddy. 2016. Project Success Prediction in Crowdfunding Environments. In *Proceedings of WSDM'16*. 247–256. https://doi.org/10.1145/2835776.2835791

[39] 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. https://doi.org/10.1287/mnsc.1120.1560

[40] Caroline Lo, Justin Cheng, and Jure Leskovec. 2017. Understanding Online Collection Growth over Time: A Case Study of Pinterest. In *Proceedings of WWW'17*. 545–554. https://doi.org/10.1145/3041021.3054189

[41] Harry Markowitz. 1952. Portfolio Selection. *The Journal of Finance* 7, 1 (1952), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

[42] Ethan Mollick. 2014. The Dynamics of Crowdfunding: An Exploratory Study. *Journal of business venturing* 29, 1 (2014), 1–16. https://doi.org/10.2139/ssrn.2088298

[43] Jagat Sastry Pudipeddi, Leman Akoglu, and Hanghang Tong. 2014. User Churn in Focused Question Answering Sites: Characterizations and Prediction. In *Proceedings of WWW'14 Companion*. 469–474. https://doi.org/10.1145/2567948.2576965

[44] PYMNTS. 2017. Crowded Crowdfunding and the Rise of Social Fundraising Fatigue. https://www.pymnts.com/platform-payments/2017/crowdfunding-social-platforms-fatigue/.

[45] Radhakrishna Rao. 1980. *Diversity and Dissimilarity Coefficients: A Unified Approach.* Technical Report 80-10. Air Force Office of Scientific Research. 38 pages. https://doi.org/10.1016/0040-5809(82)90004-1

[46] Paul Resnick, Joseph Konstan, Yan Chen, and Robert E Kraut. 2011. Starting New Online Communities. In *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, 231–280.

[47] Jeremy C. Short, David J. Ketchen, Aaron F. McKenny, Thomas H. Allison, and R. Duane Ireland. 2017. Research on Crowdfunding: Reviewing the (Very Recent) Past and Celebrating the Present. *Entrepreneurship Theory and Practice* 41, 2 (2017), 149–160. https://doi.org/10.1111/etap.12270

[48] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2018. Revisiting "the Rise and Decline" in a Population of Peer Production Projects. In *Proceedings of CHI'18*. 1–7. https://doi.org/10.1145/3173574.3173929

[49] Silvio Vismara. 2018. Information Cascades among Investors in Equity Crowdfunding. *Entrepreneurship Theory and Practice* 42, 3 (2018), 467–497. https://doi.org/10.1111/etap.12261

[50] Nir Vulkan, Thomas Åstebro, and Manuel Fernandez Sierra. 2016. Equity Crowdfunding: A New Phenomena. *Journal of Business Venturing Insights* 5 (2016), 37–49. https://doi.org/10.1016/j.jbvi.2016.02.001

[51] Rick Wash. 2013. The Value of Completing Crowdfunding Projects. In *Proceedings of ICWSM'13*.

[52] Stanley Wasserman and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge University Press.

[53] Fang Wu, Dennis M. Wilkinson, and Bernardo A. Huberman. 2009. Feedback Loops of Attention in Peer Production. (2009). arXiv:0905.1740

[54] Jiang Yang, Xiao Wei, Mark S Ackerman, and Lada A Adamic. 2010. Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities. In *Proceedings of ICWSM'10*.

[55] Juanjuan Zhang and Peng Liu. 2012. Rational Herding in Microloan Markets. *Management science* 58, 5 (2012), 892–912. https://doi.org/10.1287/mnsc.1110.1459

[56] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists through Topic Diversification. In *Proceedings of WWW'05*. 22. https://doi.org/10.1145/1060745.1060754

[57] Randi Zuckerberg. 2013. On Kickstarter Etiquette and Crowd-Funding Fatigue. https://www.huffingtonpost.com/randi-zuckerberg/kickstarter-etiquette_b_3055863.html.