Understanding Temporal Backing Patterns in Online Crowdfunding Communities

Yiming Liao
Penn State University
yiming@psu.edu

Thanh Tran
Utah State University
thanh.tran@aggiemail.usu.edu

Dongwon Lee
Penn State University
dongwon@psu.edu

Kyumin Lee
Utah State University
kyumin.lee@usu.edu

ABSTRACT

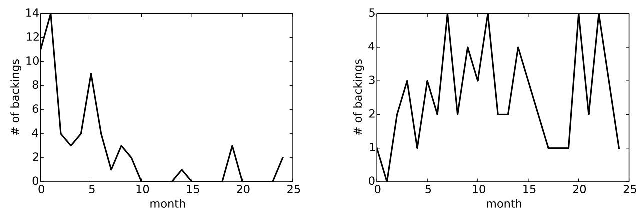
Online crowdfunding platforms such as Kickstarter and Indiegogo make it possible for users to pledge funds to help creators bring their favorite projects into life. With an increasing number of users participating in crowdfunding, researchers are progressively motivated to investigate on improving user experiences by recommending projects and predicting project outcomes. To prompt the sustainable development of these platforms, understanding backers' behaviors becomes also important, as it helps platforms provide better services and improve backer retention. In particular, studying backers' *temporal* behaviors allows them to monitor the dynamics of backers' actions and develop appropriate strategies in time. Therefore, in this paper, we analyze a large amount of backer data from Kickstarter and Indiegogo, and do a comprehensive quantitative analysis on users' temporal backing patterns. Employing time series clustering methods, we discover four distinct temporal backing patterns on both platforms. In addition, we explore various characteristics of these backing patterns and possible factors affecting backers' behaviors. Finally, we leverage these insights to build a prediction model and show promising results to identify users' backing patterns at a very early stage. The datasets used in this paper are available at: https://goo.gl/ozgLvP.

CCS CONCEPTS

• **Social and Behavior Sciences → Sociology**; • **Probability and Statistics → Time series analysis**; • **Database Applications → Data mining**;

KEYWORDS

Crowdfunding; User Behavior Analysis; Temporal Pattern



(a) From Apr. 2012 to May 2014 **(b) From Aug. 2014 to Sep. 2016**

Figure 1: Kickstarter user example

1 INTRODUCTION

Online crowdfunding platforms such as Kickstarter and Indiegogo have opened up a new avenue for entrepreneurs to raise funding, overcoming the barriers of small start-up companies [15]. In such platforms, users, known as *backers*, are given the opportunity to pledge funds to join entrepreneurs, known as *creators*, to bring their favorite projects into life. Taking Oculus Rift for example, the virtual-reality gaming headset had reportedly raised $2.4 million dollars via crowdfunding, which clearly displays the power of crowdfunding [26]. According to Kickstarter's report[1], around March 2017, 12 million backers have pledged approximately $2.9 billion dollars for more than 120k projects. Among these 12 million backers, however, only less than 4 million of backers have backed two or more projects. Similarly, less than 30% of user retention rate has been reported on other two crowdfunding platforms, DonorsChoose.org [1] and Indiegogo (Section 3).

As a motivating example, see Figure 1, where actual backing numbers of two backers on the Kickstarter platform are plotted. Note that two users show radically different backing behavior for the same duration of 24 months over different years. The user in (a) shows active backings in the early stage but gradually backs less, while the user in (b) more or less remains active throughout. Clearly, identifying such users as (a) early and encourage them to remain active is beneficial to crowdfunding platforms. To increase such backer retention rate, therefore, we argue that monitoring and understanding backers' behaviors is indispensable.

Previous studies have pointed it out that there are distinguished differences between: (1) occasional backers, who back only a small number of projects and mostly join to support their friends' projects, and (2) frequent backers, who invest on many projects and show clear backing interests [2, 22]. In addition, users' first backing behaviors are shown to be closely correlated with user retention and

WebSci'17, June 25–28, 2017, Troy, NY, USA.

ACM ISBN 978-1-4503-4896-6/17/06…$15.00.
https://doi.org/http://dx.doi.org/10.1145/3091478.3091480

[1] https://www.kickstarter.com/help/stats

even can be used to predict which backers will return [1]. Nevertheless, none of the previous crowdfunding studies has investigated on the "temporal dynamics" of users' backing behaviors over multiple years. Users' temporal backing behaviors could reflect users' dynamic relationships with platforms, which is quite valuable information for platforms to improve backer retention. For example, if a platform observes a backer gradually becomes less active, platforms may take timely actions, such as special promotion, to restore their interests. Furthermore, knowing backers' temporal behaviors enables platforms to be able to provide better personalized services, for instance, adjusting recommendation strategies at times.

Therefore, in this paper, we propose to analyze users' temporal backing behaviors on online crowdfunding platforms. In particular, we intend to investigate the following three research questions:

- **RQ1:** Do users have any temporal backing patterns?
- **RQ2:** If so, what are the characteristics of those backing patterns and possible factors impacting those patterns?
- **RQ3:** Can we identify users' backing patterns at early stages?

In answering these research questions, we make the following main contributions:

- To our best knowledge, we are the first to analyze users' temporal backing behaviors on crowdfunding platforms.
- We discover four distinct temporal backing patterns on two most popular crowdfunding platforms with hundreds of thousands frequent backers.
- We analyze the characteristics of each pattern from various aspects, including success rate and pledged money, and explore possible factors impacting users' backing patterns.
- We validate our analysis by showing the possibilities to build early prediction models in inferring all four temporal backing patterns.
- Based on the analysis, we achieve encouraging results in identifying those who will progressively become inactive at very early stages.

The rest of the paper is organized as follows. In Section 2, we review related works on crowdfunding platforms. Section 3 introduces our data collection strategy and the experimental dataset. The approach to uncover users' temporal backing behaviors and its results are given in Section 4. In Section 5, we analyze characteristics of backing patterns and explore possible factors impacting users' behaviors. Early prediction models are discussed in Section 6 and we conclude the paper with a discussion on limitations and future work in Section 7.

2 RELATED WORK

In this section, we review crowdfunding research work in three categories: (i) analysis of crowdfunding platforms, (ii) project success prediction and recommendation, and (iii) crowdfunding backer behaviors.

There are a rich set of studies on analyzing crowdfunding platforms [4, 7, 10, 11, 14, 25]. For example, studies have examined the dynamics of crowdfunding projects [19], revealed motivations of both creators and backers [11], and made comparisons of different crowdfunding platforms [12]. With the rise of social media, researchers noticed the close relationship between crowdfunding platforms and social networks, and started leveraging social media activities to study crowdfunding [2, 17, 22]. For example, social media have been found to be quite helpful in project promotion and strongly correlated with project outcomes [17]. In addition, other factors, such as geography, project updates and rewards, have also been proved to be closely connected with crowdfunding projects [16, 20, 28].

Aimed at helping creators get funded and backers find favorable projects, research has focused on using machine learning algorithms to do project success prediction [6, 9, 13, 18] and project recommendation [2, 22, 23]. Besides static and dynamic features on crowdfunding platforms, both Chung and Lee [6] and Rakesh et al. [24] leveraged social network features to achieve state-of-the-art performance in project success rate prediction and project recommendation respectively.

Although various studies have been conducted on crowdfunding platforms, few works have investigated on backers' behaviors. Previous studies have shown different backing strategies or behaviors between occasional and frequent backers. For example, frequent backers are found to be more likely to invest fast-growing projects [2]. [15] classified backers into three categories, immediate backers, delayed backers and serial backers. In addition, only based on one's first backing, researchers showed the possibility to predict whether that user will come back or not [1].

In our paper, instead of uncovering different backing behaviors between occasional and frequent backers, we intend to focus on frequent backers, who carry out more activities and are more valuable for platforms, and dig into their temporal backing behaviors, which has rarely been explored before.

3 DATASET DESCRIPTION

Aimed at understanding user behaviors in online crowdfunding communities, we collect a large amount of user data from two of the most popular crowdfunding platforms, Kickstarter and Indiegogo. To make our analysis more reliable, we have collected all projects that are available and thereby, to our best knowledge, obtain the largest datasets on both platforms (comparison of raw datasets is given in Table 1). In this section, we introduce data collection strategy, data cleaning process and overview of our dataset.

3.1 Dataset Collection

Due to the disparate strategies of Indiegogo and Kickstarter for displaying user backing histories, we adopt two different collection methods for these two platforms.

3.1.1 Kickstarter Dataset.

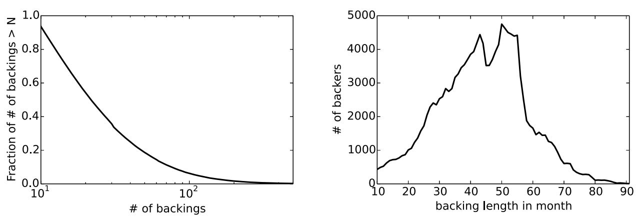
Kickstarter prevents users from obtaining backer list of certain project by only displaying limited backers' avatars without profile URLs. Fortunately, however, we can find backers' profile URLs in project comments page, and with one's profile URL, we are able to collect the complete project backing history of that user. At the end, we have gathered 233,534 ended projects, spanning from April 2009 to December 2016, and corresponding comments. Then we collected users' full backing histories through user profile URLs from comments, which resulted in more than 500,000 backers.

Table 1: Raw Dataset Comparison

Data source	Platform	# of unique users	# of projects
Ours	Kickstarter	**508,850**	**233,534**
Ours	Indiegogo	**2,314,199**	**124,292**
[6]	Kickstarter	146,721	168,851
[19]	Kickstarter	N/A	59,115
[26]	Kickstarter	239,000	N/A
[7]	Indiegogo	N/A	47,139

Table 2: Our frequent Backer Dataset

Platform	# of users	# of projects	# of backings per user
Kickstarter	150,122	174,938	38.0
Indiegogo	12,528	41,450	15.4



(a) CCDF of users' backing numbers **(b) Backing Length Distribution**

Figure 2: Kickstarter Frequent Backer Distribution

3.1.2 *Indiegogo Dataset.*

Different from Kickstarter, Indiegogo displays projects' backer lists with some anonymous backers but only shows recent backings of each user on their profile URLs. Hence, we directly collected Indiegogo's project backer lists after gathering all available ended 124,292 projects, which spans from January 2008 to January 2017. Identifying users by profile URLs, there are around 2,300,000 unique backers. Although we cannot guarantee to obtain full backing history of each backer and each user's backing history in our dataset ought to be viewed as a sample of the true history, the influence of sampling should be negligible as we have a large number of projects.

3.2 Frequent Backers

Despite a large number of unique backers identified on both platforms, the majority of them only back very limited projects. Taking Indiegogo backers for example, only 536,727 backers (23%) back more than twice, which is consistent with the previous study (26%) on another crowdfunding platform [1]. In order to reduce the impact of occasional backers and better understand users' backing behaviors, we filter out those infrequent backers and only keep those frequent backers who have backed more than 10 times. As a result, we have 150,122 Kickstarter and 12,528 Indiegogo frequent backers. Table 2 shows overall statistics of our frequent backer dataset and Figure 2(a) illustrates the complementary cumulative distribution function (CCDF) of Kickstarter users' backing numbers. For each

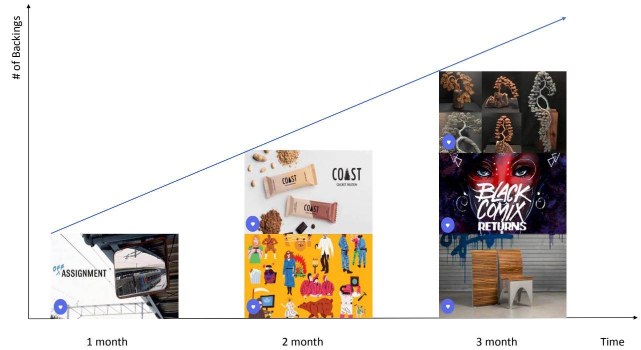


Figure 3: Example of backing behavior

backing number N, CCDF of users' backing numbers shows the proportion of backers who invest more than N projects, for example, there are more than 20% of Kickstarter backers supporting at least 40 projects.

In our paper, we analyze users' backing behaviors during their lifetime, which is defined as the observed backing history in our dataset. As shown in Figure 2(b), the majority of backers have lifetimes lying between 2 and 5 years, which means users' baking histories in our dataset are reasonably long enough for temporal backing behavior analysis.

4 TEMPORAL BACKING PATTERNS

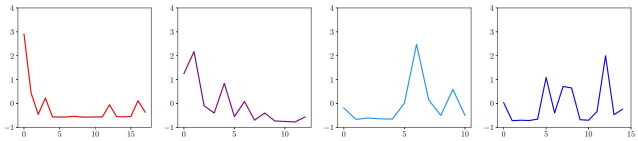
In this section, we intend to answer *RQ1: do users have any temporal backing patterns?* We start by introducing our definition of users' temporal backing behaviors, then describe our approach to cluster those temporal behaviors and finally discuss about the clustering results.

4.1 Temporal Backing behavior

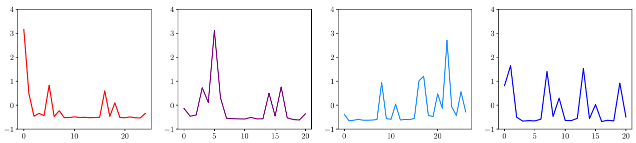
With the purpose of uncovering users' possible backing strategies, we monitor users' backing numbers over time and define a user's temporal backing behavior as a backing number function of time, $N(t)$. Although the exact timestamp when one user back a certain project is not available in our dataset, we propose to use the middle time between each project's start and end time as its backers' joining time. Since projects' average duration is around 1 month and most of our backers have more than 2 years lifetimes, it should be reasonable to ignore the several days' shifting. Knowing the timestamps of all backings, each user's backing behavior can be viewed as a time series, $N(t) = \{n_1, n_2, \ldots\}$, which models users' backing numbers through time. For example, Figure 3 shows one possible temporal backing behavior, where the user tends to back an increasing number of projects over time.

4.2 Dynamic Time Warping Clustering

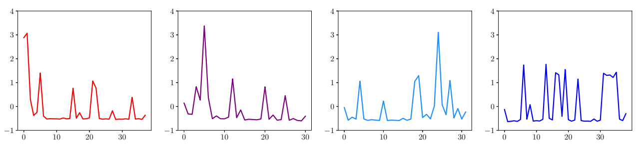
Having converted users' temporal backing behaviors into time series, we investigate whether there exist any backing patterns. Because different backing behaviors correspond to different shapes in time series, we propose to use time series clustering to find distinct patterns. Since only the shape of time series matters, we



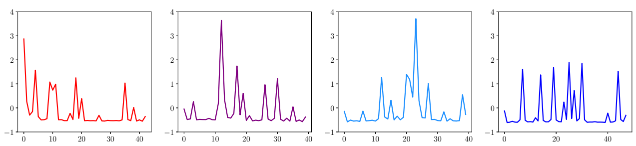
(a) 10-20 months



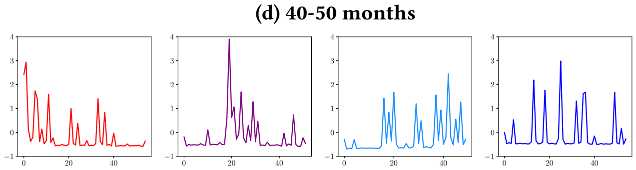
(b) 20-30 months



(c) 30-40 months



(d) 40-50 months



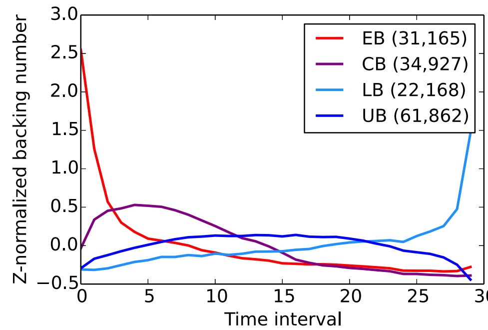
(e) 50-60 months

Figure 4: Patterns in various backing length groups. X-axis stands for months. Y-axis stands for Z-normalized backing number.

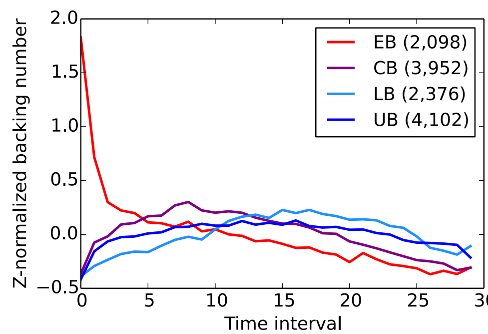
do Z-normalization on all time series at first to make comparisons between two time series meaningful [24]. Then, in order to allow slight shifting in time and warping in shape, we take advantage of Dynamic Time Warping (DTW) [5] clustering to find discriminative patterns.

More specifically, we adopt Dynamic Time Warping with Dtw Barycenter Averaging (DBA) [21], a state of the art time series averaging method, to do the clustering. As for parameter settings, we try various time window sizes and cluster numbers, and run DBA-based DTW 100 times with different kmeans++ initialization [3] for each set of parameters and keep the clusters with minimum inertia. Finally, we find the most discriminative patterns when DTW window size is $\frac{1}{3}$ of time series length and the cluster number is 4.

4.2.1 Effect of backing length.
Due to the fact that backing history length varies with users, it is difficult to compare two time series that differ too much in length (e.g. one user with 2-year backing history while another with 5-year backing history). Thus, we intend to study the effect of users' backing length by running clustering on different backing length groups



(a) Kickstarter



(b) Indiegogo

Figure 5: Clustering results (with # of backers in parentheses)

and the results are shown in Figure 4. Apparently, similar backing patterns are discovered in all groups, and there is no distinct pattern that is exclusively found in some groups. As a result, it should be reasonable to normalize each user's backing history into the same length (*Binning*): 1) we regard each user's first and last project backing time as one's lifetime start and end time respectively; 2) evenly split one's lifetime into 30 intervals and count the number of backings in each interval; 3) apply Z-normalization on the time series.

4.3 Clustering Results

After binning and clustering, we do pairwise averaging on the 30-interval time series per group to obtain smoothed clusters. As shown in Figure 5, similar backing patterns have emerged on both crowdfunding platforms, which are also consistent with clustering results on different backing length groups in Figure 4. We name the 4 distinct backing patterns as follows:

(1) **Early backer (EB)**: those backers back a lot at the very beginning of their lifetime but gradually lose interest and seldom back later.

(2) **Cautious backer (CB)**: although the majority of their backings still happen at the beginning, they are more cautious than early backers, as they try a few projects before conducting massive investments.
(3) **Late backer (LB)**: contrary to early backers, they back limited projects at first, but gradually back more.
(4) **Uniform backer (UB)**: they actively back all the time with their backings almost evenly distributed through their lifetime.

Two matters should be noted here: 1) Kickstarter clusters look much more smooth than Indiegogo clusters; 2) there is a slight difference between LB on both platforms, i.e., LB on Kickstarter shows a more sharp increase at later months than LB on Indiegogo does. The first issue results from the fact that Indiegogo dataset is smaller (with respect to # of backers) and sparser (with respect to # of backings per user) than Kickstarter dataset. As for the second issue, it may be because our Indiegogo dataset does not guarantee containing users' full backing histories and lacks some projects. However, we can still see an obvious increase in the backing pattern of LB on Indiegogo, which is matched with Kickstarter's.

5 UNDERSTANDING BACKING PATTERNS

Figure 5 indicates an interesting phenomenon that both EB and CB invest a lot at the beginning, but gradually lose interest and become less active, while UB is active in backing all the time and LB even supports an increasing number of projects over time. As EB and CB account for a large proportion (more than 40%) of frequent backers, figuring out why they have become inactive will not only help those backers out of the possible problems they are trapped in, but also contribute to the development of crowdfunding platforms. Therefore, we intend to do a quantitative analysis on the characteristics of these 4 groups of backers and propose some empirical answers to *RQ2: what are the characteristics of those backing patterns and possible factors impacting those patterns?* Note that here, we only show results on Kickstarter dataset but similar results are found on Indiegogo dataset as well, except for the success rate[2].

This section starts with analyzing characteristics of backing patterns, and based on these characteristics, we propose some possible factors explaining users' behaviors or backing patterns.

5.1 Characteristics of Patterns

When analyzing characteristics of backing patterns, cumulative distribution functions (CDF) are frequently adopted for displaying the distribution differences among those four groups in various aspects. For a certain value X, CDFs show the fraction of backers who have values less than or equal to X. To quantitatively measure the differences between distributions of two backing patterns' samples, we conduct *Kolmogorov-Smirnov* test [27], a wildly used significance test for checking whether two samples are drawn from the same distribution. Accordingly, for any pair of backing patterns in any situation, KS tests reject their samples being drawn from the same distribution with p-value less than 0.01.

5.1.1 *Backing length and backing number.*

Figure 6(a) presents the CDF of backers' lifetime backing length. As expected, EB has obvious shorter lifetime backing length than other three groups, both in general and on average, in that it gradually loses interest in backing from the beginning. Despite the fact that CB has similar backing length distribution as UB, its backing number is apparently larger than UB's as indicated in Figure 6(b), which suggests that CB is as valuable as UB. CB could even contribute more to the platforms if it remained active during the second half backing history. In addition, we examine the change of each group's backing number over time and find some inflection points. As shown in Figure 6(c)[3], on average, EB is actively backing during the first 12 months, while the passion of CB lasts 24 months, two times longer than EB's, after cautiously investing 3 months at the beginning.

5.1.2 *Project Cost.*

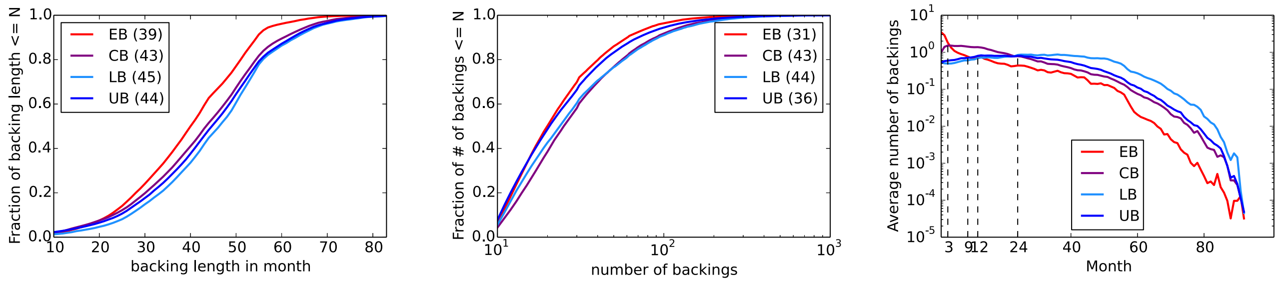
Besides participation duration and backing numbers, the amount of pledged funds is another important measurement for the contribution of users to the platforms. We compare different groups' spending behaviors and make several observations. First, as shown in Figure 7(a), on average, EB spends the least, whereas LB contributes the most. Considering the proportion of each group on the platform, UB, even if ranked 3rd on average spending, invests 151 million dollars in total, which is almost 1.5 times more than the second CB does. Because of the small number of LB, it only pledges 62 million dollars, which is slightly more than EB does. The fact that EB and CB pledge around 30% of funds among frequent backers implies that they are indispensable for the platforms. Second, Figure 7(b) demonstrates that EB and CB prefer offering smaller pledging fund per project, while LB and UB are willing to spend a little bit more on each project. Interestingly, both Figures 7(b) and 7(c) suggest EB and CB have similar average offering per project. Last but not least, Figure 7(c) shows that EB offers less money for each project when starting investments but gradually increases its offerings. We can see a similar pattern in CB's average offerings over projects as well. Taking together with their temporal backing patterns, this phenomenon may occur due to their limited budget. At the beginning, in order to support multiple projects, EB and CB have to lower down their offering per project, but with their backing numbers gradually decreasing, they are able to afford more offering per project.

5.1.3 *Success Rate.*

Next, we investigate on the success rate of each backing group. Surprisingly, 22.4% of the 150K frequent backers (with at least 10 backings) in Kickstarter dataset have never failed in backing, even though their average backing number is 18. As for each backing group, 25.1% of UB, 24.4% of LB, 21.2% of EB and 17.6% of CB achieve the perfect backing histories. Although we analyze frequent backers, who have up to 90.5% success rate on average, there are still some differences among these four groups. Figure 8(a) shows that LB and UB are better at backing and have higher success rate than other two groups, which is consistent with the ratios of perfect backers in each group to some extent. Moreover, looking at each group's

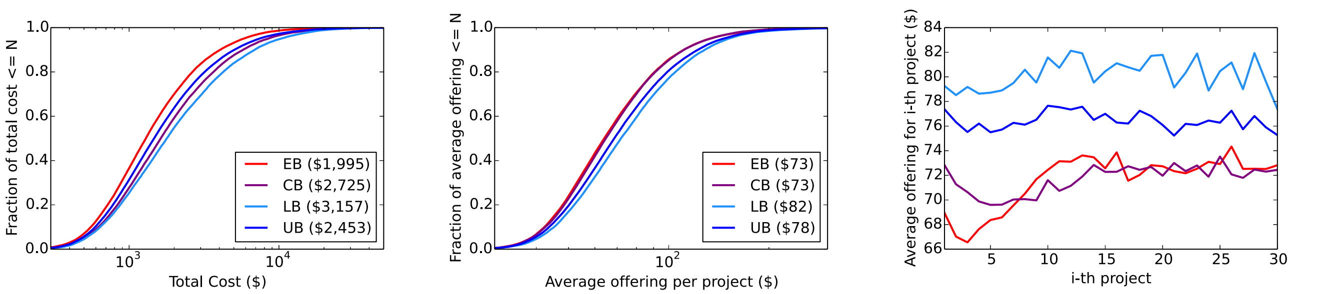
[2] Besides *All-or-Nothing*, Indiegogo gives one more choice, *Take-it-All*, to creators, which means backers are still able to get their rewards even if the goal amount is not reached. With more than 90% of Indiegogo projects starting with flexible funding, we cannot determine whether projects are successful or not on Indiegogo dataset.

[3] Note that it is the overall trend that matters not the absolute values, which are relatively small due to the data sparsity (not all users invest projects in every month).



(a) CDF of backing history length **(b) CDF of total backing number** **(c) Average backings per month.**

Figure 6: Backing length and backing number of each pattern (with average value in parentheses)



(a) CDF of total cost **(b) CDF of average offering** **(c) Average offering for each backer's i-th project**

Figure 7: Backing cost (with average value in parentheses). Cost refers to the money spent on successful projects, while offering just indicates the prices that users are willing to offer for the projects, which may or may not be successful.

average success rate at i-th project from Figure 8(b), we can see a clear gap between success rates of EB and CB and those of LB and UB. In addition, we note the slopes of CB and EB's curves are relatively smaller, indicating a faster drop in their success rates. Considering temporal backing patterns of EB, the low success rate may result from the fact that EB invests very frequently at the beginning without caring much about its backings' outcomes, while other three groups, especially LB, gain more experience before conducting massive investments and certainly reach much higher success rate in backings.

5.1.4 Categories and Creators.
Previous studies have shown that topical preference and connections with creator play an important role in users' backing behaviors [11, 22]. Category entropy is applied to measure users' topical preference and defined as follows:

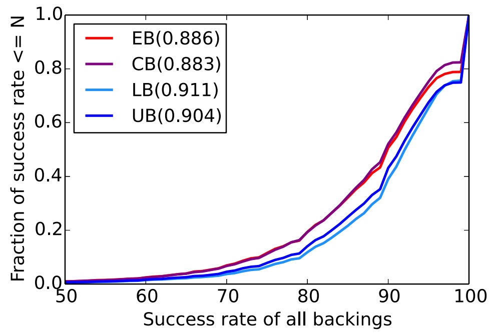
$$CatEntropy(u) = -\sum_{i=1}^{C} \frac{n_i}{N} log_2 \frac{n_i}{N} \tag{1}$$

In the equation, C is the number of categories, N is the total number of projects backed by user u and n_i is the number of backed projects under category i. Compared with uniformly backing all 15 project categories on Kickstarter, where category entropy is around 4, Figure 9(a) indicates that all four groups have strong preferences in backing categories, which is consistent with previous studies. We also notice that EB and CB have larger entropies than other two groups in general, which means they have broader interests in backing projects. As expected, Figure 9(b) shows that there are big chances that frequent backers will support creators whom they have supported before. In addition, the probability of LB and UB funding a past supported creator is at least 40% higher than that of EB and CB during the first 20 backings. Although EB and CB's lower probability at the first a few backings may result from their more frequent backings in a short time at the beginning, they are still less likely to fund a past supported creator than LB and UB after a longer time, say after 20 backings.

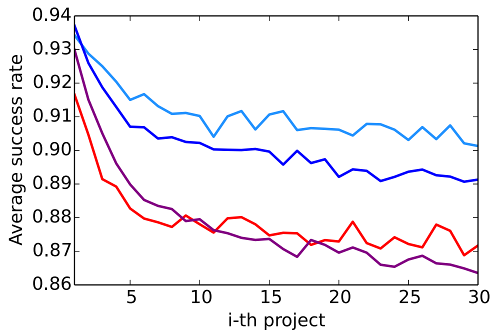
5.2 Causes of Patterns

Having observed various characteristics of users' temporal backing patterns, we seek potential factors impacting users' behaviors. Since users' backing behaviors can rely on diverse factors, including personal backing strategies, we just provide some general empirical analysis here and leave user studies for future investigation. Specifically, we focus on the impact of project outcomes and creators in this section.

5.2.1 Project Success.
In Section 5.1.3, we have identified the huge differences among each group's backing success rate that success rates of EB and CB are not only much lower than those of LB and UB, but also drop faster than other two groups at the early stage. Considering EB and CB gradually lose interests in backing from the beginning, the



(a) CDF of success rate



(b) Average success rate for each backer's i-th project

Figure 8: Success rate (with average value in parentheses)

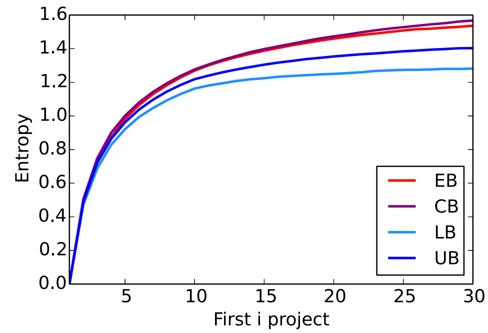
relatively low success rate should play an important role in their later inactivity. Thus we argue that EB and CB may be discouraged by experiencing relatively low backing success rates and failing several projects initially, and thereby progressively become less active in backing. Based on this finding, in order to encourage users to be active, platforms may recommend some easily successful projects to those who have a bad start initially in backing.

5.2.2 Connections with Creators.

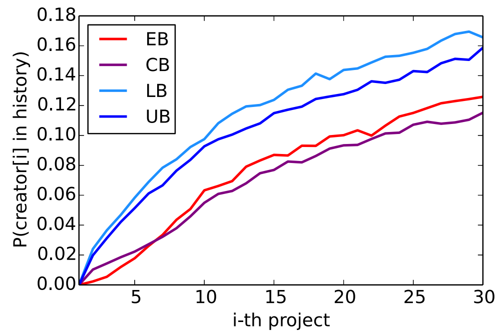
Another interesting observation comes from Section 5.1.4. Note that both LB and UB are more likely to fund their past supported creators, indicating stronger connections between them and their creators. During the first 20 backings, after which EB and CB gradually become inactive, LB and UB show at least 40% higher probability to support a recognized creator. Because of LB and UB's willingness to return to fund past supported creators, there could be some long-term connections being established between them and their creators. Accordingly, favorable relationship between users and creators may increase the probabilities of users being a returned backer, and encourage users to continue backing on the platforms. Hence, both platforms and creators should pay more attention to building up mutual trust between backers and creators.

6 EARLY PREDICTION

In addition to the differences in the dynamics of backing numbers, we have shown clear distinctions among these 4 groups in various



(a) Average entropy of each backer's first i projects



(b) Probability of each backer's i-th project being created by his/her past supported creator

Figure 9: Categories and creators

Table 3: Duration in days at i-th finished project

at i-th finished project	1	2	3	4	5	6
Kickstarter	18	104	166	220	271	320
Indiegogo	26	162	252	327	396	465

aspects, including project cost and success rate, in the previous section. To validate the insights from the previous section, we build up models on users' early backing features to see whether we are able to predict users' later backing patterns. Apparently, the earlier we can detect the loss of users' interests, the better we may help them stay active by taking actions such as promotion. Therefore, we turn to *RQ3: can we identify users' backing patterns at early stages?* and investigate on building early prediction models in this section.

6.1 Set-up

To validate our analysis, we build 4-class classifiers to predict users' backing patterns at first. Then, based on the observation in section 5.1.2 that EB and CB pledge around 30% of funds among frequent backers and had the potential to contribute even more if they stayed active, we propose to build binary classifiers to distinguish EB and CB from UB and LB as well.

6.1.1 Prediction Time.

Section 5.2.1 points out the influence of project outcomes on users' future behaviors. As such, it should be reasonable to make inferences after observing a few project outcomes. Table 3 displays Kickstarter and Indiegogo users' duration in days[4] when their i-th project is finished. As shown in Figure 6(c), since 9th month, EB has already been less active than all other three groups, which might be too late for user retention. Consequently, we propose to predict users' backing patterns at the end of their 1st to 4th project on Kickstarter and correspondingly, at the end of 1st to 3rd project on Indiegogo.

6.1.2 Features.

After setting up the prediction time, we extract the following features from users' behavior data during the time period to build classifiers. Note that, when clustering users' backing patterns, we only used Z-normalized time series of users' backing numbers in their entire lifetimes. However, here we only utilize users' first a few months data to extract features.

(1) **Temporal (4 features)**: As indicated in Figure 6(c), even when just joining the platforms, different groups of users behave differently in backing. Therefore, dynamics of users' backing numbers should be valuable features. In addition, we include the number of backed projects[5], duration in days, mean and standard deviation of the time difference between two adjacent backings.

(2) **Backer (3 features)**: In sections 5.1.2 and 5.1.3, we observe that averaging offering per project and success rate vary with different groups in the first a few backings. Therefore, we formulate backer features, including backing success rate, mean and standard deviation of backed money per project.

(3) **Creator (4 features)**: Sections 5.1.4 and 5.2.2 discuss about the potential encouragement of favorable connections between backers and creators. Thus, based on the assumption that experienced creators may be good at establishing favorable relationship with backers, we also extract features from creators, involving their past backing numbers, backing success rate, creating numbers, and creating success rate.

6.1.3 Criteria.

Because of the imbalance of our data, we evaluate the multi-class classification results using two popular measures–i.e., macro-F1 score and G-means, to be defined as follows:

$$Macro\text{-}F1 = \frac{1}{N_c} * \sum_{i=1}^{N_c} F1_i \tag{2}$$

$$G\text{-}means = \sqrt[N_c]{\prod_{i=1}^{N_c} recall_i} \tag{3}$$

with

$$F1_i = \frac{2 * precision_i * recall_i}{precision_i + recall_i}$$

[4]By the end of i-th project, how long have that user been on the platform?

[5]At the end of i-th project, users may back more than i projects, though their outcomes are not available then.

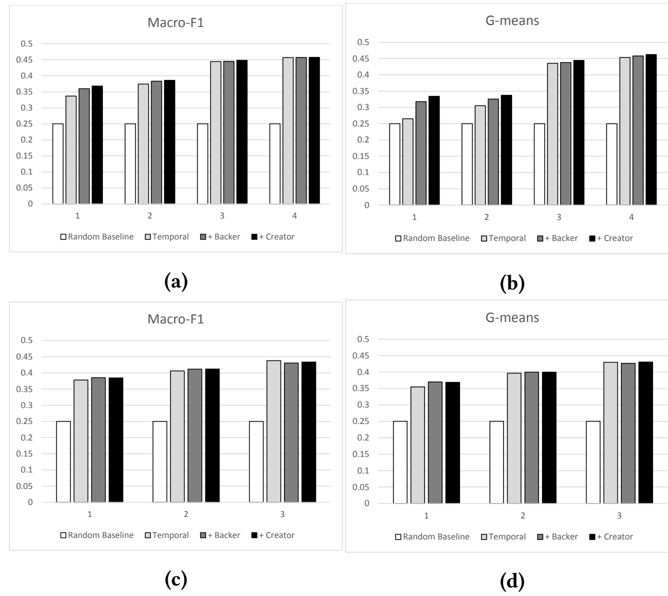


Figure 10: Predictive performance for inferring users' future behavior patterns at the end of first few projects. (a) and (b) show results on Kickstarter dataset, while (c) and (d) display Indiegogo's.

$$precision_i = \frac{\sum TP_i}{\sum TP_i + \sum FP_i}$$

$$recall_i = \frac{\sum TP_i}{\sum TP_i + \sum FN_i}$$

In the equations, N_c stands for the number of classes, TP for True Positive, FP for False Positive, and FN for False Negative. For 4-class classification, random baselines will score 0.25 on both metrics.

As for binary classification, Area Under the ROC Curve (AUC) is employed to measure the possibility that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. Thus, a random baseline will have 0.50 of AUC score.

To handle our imbalanced data, via preliminary testing, we chose the AdaBoost algorithm [29] with a decision tree classifier and under-sampling [8], which has given the best performance on 10-fold cross validation.

6.2 Results

6.2.1 Predicting backing patterns.

The results are given in Figure 10. Generally, based on our proposed features, we can achieve promising performance on inferring users' future backing patterns when they just complete first 3 to 4 projects. As expected, backer and creator features contribute to the prediction, especially at the end of their first or second project, which validates our analysis in the previous section. With more user backings being observed, although overall performance is improved significantly, temporal features play a more essential role in the prediction and other two features only lead to smaller increments. Consistent with the fact that success rate is indeterminable on Indiegogo dataset as mentioned in section 5, backer and creator features show very limited improvement in the prediction.

Table 4: Binary classification results (ROC AUC)

at i-th finished project	1	2	3	4
Random Baseline	0.50	0.50	0.50	0.50
Kickstarter	0.71	0.78	0.81	0.82
Indiegogo	0.63	0.70	0.73	0.75

6.2.2 *Identifying early backers and cautious backers.*

Having validated our analysis and corresponding features, we build a binary classifier to identify EB and CB from all frequent backers for user retention. As shown in Table 4, at the end of each user's 1st project, we are able to tell whether that user will be EB and CB with 0.71 AUC on Kickstarter and 0.63 AUC on Indiegogo. When one's finished project number reaches 4, AUC increases to 0.82 and 0.75 respectively. Additionally, when joining the platforms for just 5 months, which corresponds to the time of 3rd project finished on Kickstarter and second project finished on Indiegogo, EB and CB can be identified very well by the built classifier, with 0.81 AUC and 0.70 AUC, correspondingly.

6.2.3 *Discussion.*

Although 4-class classification models do not achieve superior performance in backing pattern inference, we show the possibility of building predictive models, which confirms the relationship between users' early activities and their future behaviors. Based on the assumption that platforms may do some user retention to help potential EB and CB stay active, we formulate the identification of EB and CB as a binary classification problem and obtain very promising results. At the end of users' 3rd project, we are able to identify EB and CB with 0.81 AUC on Kickstarter and 0.73 AUC on Indiegogo. Certainly, there is still much room for improvement in EB and CB identification. For example, we can incorporate more user behavioral features, such as users' view and clicking histories. We leave this for future study. Throughout these preliminary studies, we emphasize that there are four distinct backing patterns, which are proved to be strongly related with users' very early behaviors, and show encouraging results in identifying those who tend to progressively become inactive at early stages.

7 CONCLUSION

7.1 Concluding Remarks

Having a better understanding of the backers' dynamic and temporal behaviors allows crowdfunding platforms to provide better services and improve customer retention. In spite of such an importance, however, prior studies lack the exploration of the temporal dynamics of backer behaviors. As such, this paper aimed to take a step to investigate on backers' temporal behaviors.

To analyze backers' temporal behaviors, we have collected large-scale datasets from two of the most popular crowdfunding platforms, Kickstarter and Indiegogo. Employing time series clustering methods, we discovered four distinct temporal backing patterns on both platforms (**RQ1**), including Early Backer (EB), Cautious Backer (CB), Uniform Backer (UB), and Late Backer (LB). Among these patterns, both EB and CB show decreasing interests in backing, whereas UB constantly invests on projects and LB even progressively supports an increasing number of projects. Driven by the research question why backers have such temporal backing patterns (**RQ2**), we conducted a quantitative analysis, explored the characteristics of these patterns in various aspects, and proposed possible factors affecting users' backing behaviors.

In our findings, we noted that the outcomes of backers' first a few projects are closely correlated with their future backing behaviors. For instance, initial low success rate may discourage users from investing further. In addition, we found that the connections with project creators could be an another factor impacting users' backing behaviors–that is, favorable relationship between backers and creators seems to encourage backers to keep supporting the same creators and/or invest on more projects. In order to validate our analysis, we extracted related features from users' first few backings, and showed promising results in early predicting all four temporal backing patterns (**RQ3**). Moreover, aimed at helping platforms with respect to user retention, we proposed to build binary classification models to identify two types of backers–EB and CB–at a very early stage and achieved encouraging performance.

7.2 Limitations and Future Work

Despite encouraging results, our work is not without limitations.

First, as usual for data-driven study using social media data, our study is based on a small fraction of real-life datasets sampled from two crowdfunding platforms. Due to the limitation of programming APIs, we had to adopt two different data collection strategies that may have introduced two types of sampling bias: (1) For Kickstarter dataset, as we obtained users' profile URLs from projects' comments, backing histories of those backers who seldom leave comments are missing. However, we hypothesized that frequent backers care more about their backings than occasional backers and therefore should be more active in leaving comments. (2) On Indiegogo, in each project's backer list, there are lots of anonymous users that we could not identify. Therefore, our dataset might be from those users who are less concerned about their privacy. In addition, we could not obtain users' full backing histories, which may affect some users' temporal backing patterns and make them act as noises in our dataset. As to these potential issues in our datasets, we attempted to mitigate the effect by obtaining relatively large-scale datasets and verified that there were consistent results on both platforms. Nevertheless, to be able to generalize our findings further, we plan to collect less-biased datasets and repeat our study on other types of crowdfunding platforms (e.g., Change.org) as well.

Second, in this paper, we mainly focused on the relationship between users' temporal backing behaviors and project features. However, it is worth noting that user features, including users' demographics and logging histories, could be equally informative. Therefore, we intend to explore connections between users' temporal backing behaviors and user features.

Third, although our investigation was based on quantitative analysis, some unforeseen factors affecting users' backing behaviors may not be reflected by the collected data. For example, users' backing behaviors might be influenced by their friends'. Accordingly, conducting user studies can be a good complement to fully understand users' temporal behaviors. Such a study will not only validate our findings in this paper but also provide potential causes of backers' actions.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions. This work was supported in part by NSF CNS-1553035, NSF CNS-1422215, NSF IUSE-1525601, and Samsung GRO 2015 awards. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

REFERENCES

[1] Tim Althoff and Jure Leskovec. 2015. Donor retention in online crowdfunding communities: A case study of donorschoose. org. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 34–44.

[2] Jisun An, Daniele Quercia, and Jon Crowcroft. 2014. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd international conference on World wide web*. ACM, 261–270.

[3] David Arthur and Sergei Vassilvitskii. 2007. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1027–1035.

[4] Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher. 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing* 29, 5 (2014), 585–609.

[5] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.

[6] Jinwook Chung and Kyumin 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*. ACM, 211–220.

[7] Douglas J Cumming, Gaël Leboeuf, and Armin Schwienbacher. 2014. Crowdfunding models: Keep-it-all vs. all-or-nothing. In *Paris December 2014 finance meeting EUROFIDAI-AFFI paper*, Vol. 10.

[8] Chris Drummond, Robert C Holte, et al. 2003. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on learning from imbalanced datasets II*, Vol. 11. Citeseer.

[9] 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*. ACM, 177–182.

[10] Elizabeth M Gerber and Julie Hui. 2013. Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 34.

[11] Elizabeth M Gerber, Julie S Hui, and Pei-Yi Kuo. 2012. Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. In *Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics*, Vol. 2. 11.

[12] Michael D Greenberg, Julie Hui, and Elizabeth Gerber. 2013. Crowdfunding: a resource exchange perspective. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 883–888.

[13] 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*. ACM, 1815–1820.

[14] 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*. ACM, 62–74.

[15] Venkat Kuppuswamy and Barry L Bayus. 2015. Crowdfunding creative ideas: The dynamics of project backers in Kickstarter. (2015).

[16] Yusan Lin, Wang-Chien Lee, and Chung-Chou H Chang. 2016. Analysis of rewards on reward-based crowdfunding platforms. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE, 501–504.

[17] 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*. ACM, 573–582.

[18] 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*. ACM, 49–61.

[19] Ethan Mollick. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of business venturing* 29, 1 (2014), 1–16.

[20] Michael Muller, Werner Geyer, Todd Soule, and John Wafer. 2014. Geographical and organizational distances in enterprise crowdfunding. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 778–789.

[21] François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44, 3 (2011), 678–693.

[22] Vineeth Rakesh, Jaegul Choo, and Chandan K Reddy. 2015. Project Recommendation Using Heterogeneous Traits in Crowdfunding.. In *ICWSM*. 337–346.

[23] Vineeth Rakesh, Wang-Chien Lee, and Chandan K Reddy. 2016. Probabilistic Group Recommendation Model for Crowdfunding Domains. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 257–266.

[24] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 262–270.

[25] Michael Siering, Jascha-Alexander Koch, and Amit V Deokar. 2016. Detecting Fraudulent Behavior on Crowdfunding Platforms: The Role of Linguistic and Content-Based Cues in Static and Dynamic Contexts. *Journal of Management Information Systems* 33, 2 (2016), 421–455.

[26] Michael A Stanko and David H Henard. 2016. How Crowdfunding Influences Innovation. *MIT Sloan Management Review* 57, 3 (2016), 15.

[27] R Wilcox. 2005. Kolmogorov–smirnov test. *Encyclopedia of biostatistics* (2005).

[28] 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*. ACM, 591–600.

[29] Ji Zhu, Hui Zou, Saharon Rosset, and Trevor Hastie. 2009. Multi-class adaboost. *Statistics and its Interface* 2, 3 (2009), 349–360.