5-2016

Understanding the Characteristics of Successful Projects and Post-Campaign Activities in a Crowdfunding Platform

Madhavi Reddy Dontham

Utah State University

UNDERSTANDING THE CHARACTERISTICS OF SUCCESSFUL PROJECTS AND POST-CAMPAIGN ACTIVITIES IN A CROWDFUNDING PLATFORM

by

Madhavi Reddy Dontham

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Electrical Engineering

Approved:

Kyumin Lee Major Professor	Nicholas Flann Committee Member
Haitao Wang Committee Member	Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2016

ABSTRACT

Understanding the Characteristics of Successful Projects and Post-Campaign Activities in a Crowdfunding Platform

by

Madhavi Reddy Dontham, Master of Science

Utah State University, 2016

Major Professor: Kyumin Lee
Department: Electrical and Computer Engineering

Online crowdfunding platforms provide project creators with new opportunities for seeking funds from people in the world. But reaching a fund-raising goal or making a project successful is always a challenge. Besides, little is known about post-campaign activities of project creators and backers. To fill the gap, in this research, we are interested in understanding (i) the characteristics of successful projects, (ii) how project creators reacted when their projects failed, and (iii) what post-campaign activities creators and backers made. To achieve our research objectives, first, we analyzed successful projects and failed projects on Kickstarter, the most popular crowdfunding platform. Then we clustered successful projects by their evolutionary patterns in terms of pledged money toward understanding what efforts project creators should make in order to make a project successful and get more pledged money. We also analyzed what activities project creators and backers made during a post-campaign period by building topic models from comments associated with the projects.

(42 pages)

PUBLIC ABSTRACT

Understanding the Characteristics of Successful Projects and Post-Campaign Activities in a Crowdfunding Platform

Madhavi Reddy Dontham

Online crowdfunding platforms provide project creators with new opportunities for seeking funds from people in the world. But reaching a fund-raising goal or making a project successful is always a challenge. Besides, little is known about post-campaign activities of project creators and backers. To fill the gap, in this research, we are interested in understanding (i) the characteristics of successful projects, (ii) how project creators reacted when their projects failed, and (iii) what post-campaign activities creators and backers made. To achieve our research objectives, first, we analyzed successful projects and failed projects on Kickstarter, the most popular crowdfunding platform. Then we clustered successful projects by their evolutionary patterns in terms of pledged money toward understanding what efforts project creators should make in order to make a project successful and get more pledged money. We also analyzed what activities project creators and backers made during a post-campaign period by building topic models from comments associated with the projects.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Crowdfunding, or crowd-sourced fund-raising, provides a revolutionary way to support ideas and projects across a number of industries, such as technology, music, film and art [1]. From reward-based crowdfunding platforms like Kickstarter, Indiegogo, and RocketHub, to donation-based crowdfunding platforms like GoFundMe and GiveForward, to equity-based crowdfunding platforms like CrowdCube, EarlyShares and Seedrs - these platforms have shown the effectiveness of funding projects from millions of individual users [2]. A Crowdfunding industry report [3] stated that a World Bank-commissioned study that hypothesizes that global crowdfunding could grow to $96 Billion by 2020. This explains that crowdfunding has drawn substantial attention from various sectors of business.

The crowdfunding process has mainly two subjects: 1) creator, who proposes the innovative idea and sets the goal and duration of the project 2) backer, who supports the project by investing smaller to huge funds on the project. Every project has to reach the predefined Fund-raising goal within the proposed duration. Even though crowdfunding has drawn significant attention globally, there is a very little examination of (successful and failure) projects in research communities. Specifically, there is no investigation on the reactions of creators after project failures and about there is no knowledge on the activities of projects after successful campaigns. Thereby, we first analyze the behavior of fund-raising activities of successful projects to understand how these projects are different and revealing what strategy project creators should use to increase pledged money. Secondly, we investigate the activity of projects after successful campaigns and analyze the post campaign behavior to answer following research questions: how active were the backers and creators after the campaign is successful? Did all the creators deliver rewards to the backers within the mentioned estimated delivery date? What did backers and creators talk about after fund raising campaign? With the increase in the number of projects and amount of pledged funds on crowdfunding platforms, the success rate of projects has been decreasing and resulting more number of failure projects. Hence, we analyze how project creators behaved after their projects failed to answer the following questions: Did they give up and no longer create projects? Or did they continue to create projects? If they continued creating projects with the same idea of the failed projects, what changes did they make in order to make the projects successful?

Towards answering these questions, we made the following contributions in this research:

- We used the largest dataset, consisting of all Kickstarter project pages, user pages, each project's temporal data and each user's Twitter account information.
- We clustered successful projects towards understanding how these clusters are different and revealing what strategy project creators should use to increase pledged money.
- We analyzed what reactions project creators had when the project failed. If they relaunched the failed projects with some improvements and made them successful, what efforts they would make.
- We build the ground truth for analyzing the behavior of successful projects after the campaign.
- We performed topic modeling on the comments to understand what topics interested the backers most.
- We have analyzed distinguishing features of creators and backers and also project traits in on-time delivered projects and late delivered projects.

CHAPTER 2
RELATED WORK

In this section we discuss the crowdfunding research carried out specifically on two categories: 1) analysis of crowdfunding platforms, fund-raising campaign and post-campaign activities; and 2) classification of backers and projects.

The crowdfunding platforms have been analyzed by many researchers [4–7]. Kuppuswamy [8] studied about backer dynamics over the project funding cycle. Mollick [9] examined the dynamics of success and failure among crowdfunded ventures. Joenssen and Müllerleile [10] analyzed 42,996 Indiegogo projects, and found that scarcity management was problematic at best and reduced the chances of projects to successfully achieve their target funding. Althoff and Leskovec [11] presented various factors impacting investor's retention and identified various types of investors. The researchers found that investors are more likely to return if they had a positive interaction with the receiver of the funds. Etter et al. [12] have studied that the time series of money pledges to classify campaigns as probable success or failure reached a higher accuracy. Authors in [13] have build a feedback tool that would be the comparison of the traits of an individuals project to the traits of other, successful projects, in a manner similar to the research through design approach pioneered in HCI [14].

Researchers also studied about activities during the project campaigns. Lu et al. [15] analyzed the hidden connections between the fund-raising results of projects on crowdfunding websites and the corresponding promotion campaigns in social media. An et al. [16] proposed different ways of recommending investors by using hypothesis-driven analyses. Naroditskiy et al. [17] investigated whether viral marketing with incentive mechanisms would increase the marketing and found that providing a high level of incentives resulted in a statistically significant increase. Jacob, in [18], simulated the dynamics of crowdfunding site to examine how the presence of high-performing superstar projects on a crowdfunding site affects donors ability to coordinate their actions and fund other less popular but still worthwhile projects on the site. Mitra in [19] have found that the language used for the project has greater predictive power accounting for 58.56% of the variance around successful funding. The analysis in [20]showed that particular uses of updates had strong association with campaign success compared to the projects description

Other researchers have classified backers and projects to various types. Kuppuswamy and Bayus [8] classified backers into three categories – immediate backers, delayed backers and serial backers. Hemer [21] classified crowdfunding projects into for-profit or not-for-profit projects. Haas et al. [22] also classified projects into hedonistic or altruistic projects using a clustering algorithm from a business standpoint.

Compared with the previous research work, we collected the largest datasets consisting of all Kickstarter project pages, corresponding user pages, each project's temporal data , and conducted a comprehensive analysis of activities in crowdfunding platforms. To our knowledge, we are the first to study to analyze the failure projects and what efforts project creators made for the later success of the projects. We have used a Gaussian mixture model-based clustering algorithm, we clustered successful projects to understand how these clusters were different and how project creators increase pledged money. Finally, we analyze the projects after the successful campaign and perform LDA topic modeling of comments to understand the interesting topics among backers and creator's discussion on the Kickstarter platform. We also analyzed the activeness of backers and creators in post-campaign.

CHAPTER 3

DATASET

In this chapter we discussed about the dataset we used for the proposed research. To analyze the fund-raising campaign and post-campaign activities on crowdfunding platforms and also to understand the characteristics of successful and failure projects, we have used the data collected by Chung and Lee [2], from Kickstarter, the most popular crowdfunding platform.

Static data: We have used 168,851 project pages which were created between 2009(Kickstarter site was launched in 2009) and September 2014. A project page consists of a project duration, funding goal, project description, rewards description and many other features. We also collected corresponding 146,721 distinct user pages each of which consists of bio, account longevity, location information, the number of backed projects, the number of created projects, and so on. Among 168,851 project pages, we filtered 17,243 projects which have been either canceled or suspended, or in which the project creator's account has been canceled or suspended. Among 146,721 user pages, we filtered corresponding 14,435 user pages. Finally, 151,608 project pages and 132,286 user pages have been used for the research.

Time series Data. To understand the patterns of successful projects, we clustered the projects and analyzed these clusters to understand how external activities affect project's temporal patterns. We need temporal data for this analysis and hence we used temporal data of 74,053 projects which were created between March 2013 and August 2014 and were ended by September 2014.

CHAPTER 4

CLUSTERING SUCCESSFUL PROJECTS AND ANALYZING THE CLUSTERS

In this section, we aim to (i) cluster successful projects based on a time series of normalized daily pledged money, (ii) analyze what kind of clusters we find and how the clusters are different from each other and understand (iii) how external activities affected projects temporal patterns.

4.1 Preprocessing Data

Out of 74,053 projects containing temporal data, we selected successful projects each of which had a project goal equal to or greater than \$100 since it is less interesting to find patterns from projects whose goal is less than \$100, considering them as noisy data. Finally, we had 30,333 successful projects. Since each project has the different duration (e.g., 30 days or 60 days), first, we converted each project duration to 20 states (time slots). Then, in each state, we measured obtained pledged money during each state. We created 20 temporal/time buckets and inserted each project's pledged money during each state to each bucket (e.g., the 1st bucket contains each project's pledged money obtained during the first state – first 5% duration in this context). To make sure which project got relatively higher or lower pledged money in each bucket, first we measured the mean (μ) and standard deviation (σ) of pledged money in each bucket for all successful projects. Then, we normalized pledged money (pm_i) of each project in the ith bucket (i.e., pledged money obtained during the ith state) as follows:

$$p\bar{m}_i = \frac{pm_i - \mu_i}{\sigma_i}$$

where μ_i and σ_i are the mean and standard deviation of pledged money of the successful projects in a ith bucket.

After running the normalization in each bucket for the projects, we had a time series of relative pledged money for each project and used these time series in the following subsections.

4.2 Clustering approach

To identify clusters of successful projects, we applied Gaussian Mixture Model (GMM) based clustering algorithm. GMM based clustering approach has been widely used by other researchers in other domains such as clustering experts in a question-answering community [23] and image processing [24, 25].

We formally define our clustering problem as follows: Given vectors $X = \{x_1, x_2, ..., x_N\}$ of N independent projects, where x_i represents a time series vector of relative pledged money in ith project, we applied GMM based clustering algorithm to find K clusters amongst observed N time series in X.

By using GMM, the log likelihood of the observed N time series is written as follows:

$$lnP(X \mid \pi, \mu, \Sigma) = \sum_{i=1}^{N} ln\Bigg\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \Bigg\}$$

, where the parameter $\{\pi_k\}$ is the mixing coefficients of a cluster k and must satisfy two conditions: $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$. μ_k and Σ_k are the mean and covariance matrix of the cluster k, respectively. $\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$ is the multivariate Gaussian distribution of cluster k, defined as follows:

$$\mathcal{N}(x_i \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{\mid \Sigma_k \mid^{1/2}} exp\Bigg\{ -\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \Bigg\}$$

We used EM algorithm to maximize the log likelihood function with regard to parameters including means μ_k, covariance Σ_k and the mixing coefficient π_k. We first initialized the values of these parameters. Then in Expectation step, the responsibilities $\gamma_k(x_i)$ of the k^{th} component of observation x_i was calculated by the current parameter values with regard to Bayesian theorem as follows:

$$\gamma_k(x_i) = p(k|x_i) = \frac{p(x_i)p(x_i|k)}{\sum_{l=1}^{K} p(l)p(x_i|l)} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}$$

In Maximization step, parameters μ_k, Σ_k and π_k were re-estimated by using the current responsibilities as follows:

$$\mu_k^{new} = \frac{1}{\sum_{i=1}^{N} \gamma_k(x_i)} \sum_{n=1}^{N} \gamma_k(x_i) x_i$$

$$\Sigma_k^{new} = \frac{1}{\sum_{i=1}^{N} \gamma_k(x_i)} \sum_{n=1}^{N} \gamma_k(x_i)(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{\sum_{i=1}^{N} \gamma_k(x_i)}{N}$$

Then, the log likelihood was evaluated. The EM algorithm was stopped when the convergence condition of log likelihood was satisfied or the number of iterations exceeded a pre-defined value.

To estimate the optimal number of clusters inputting in GMM, we used the Bayesian Information Criteria (BIC). In statistics, BIC is a criterion based on the likelihood function for model selection among a finite set of models. The model with the lowest BIC value is the best one among the models. In our study, a model with the lowest BIC value indicates that the number of clusters K in the model is the optimal number, returning the most meaningful clusters. Let $\widehat{L}$ as the maximum value of the likelihood function of the model, the value of BIC is calculated as following:

$$BIC(K) = -2ln\widehat{L} + K \ln N$$

4.3 Analysis of clusters

In this section, we try to understand the various patterns of successful projects and also study the impact of project features on final pledged funds in every cluster.

To find the optimal number of clusters, we ran the GMM based clustering algorithm in a range of $K = 1 \sim 20$ by increasing 1 in each time, and got a BIC value in each case. Figure 4.1 depicts a BIC curve showing how a BIC value was changed as we increased K by

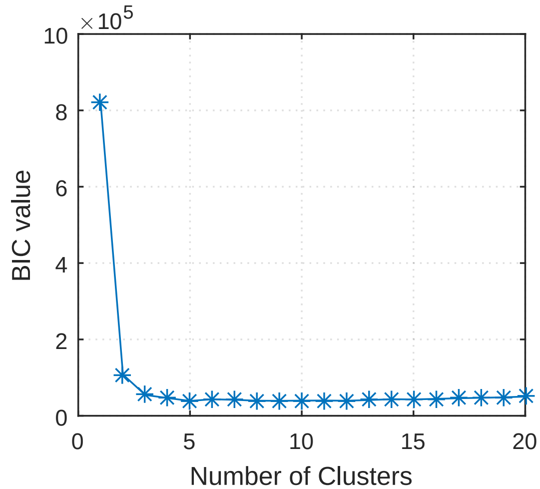


Fig. 4.1: BIC curve for successful projects

1 in each time. Finally, $K = 5$ returned the smallest BIC value and returned the optimal 5 clusters.

To understand how each cluster had different temporal patterns, we measured the mean of relative pledged money in each bucket of projects in each cluster. Then, we drew a line of the means for each of the five clusters of successful projects as shown in Figure 4.2.

- Projects in a cluster C2 received an almost same amount of relative pledged money over time.

- Projects in a cluster C3 received the largest amount of pledged money over time

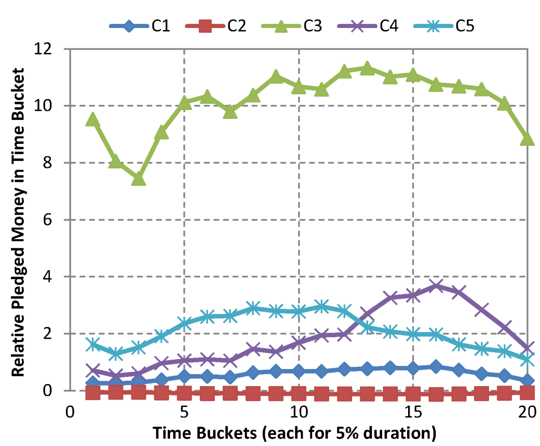


Fig. 4.2: Evolutional patterns of five clusters.

compared with projects in the other four clusters. In the beginning, relative pledged money went down until the 3rd time bucket, went up until the 13th time bucket with some fluctuation, and then gradually went down. Why did this evolutional pattern happen? We conjecture that the news of initial popularity was propagated to other users, some of whom eventually backed up the projects, increasing daily/relative pledged money. It is a typical evolutional pattern of the most popular projects like the Coolest Cooler and the Pono Music [1].

- A cluster C4 had the most interesting pattern. The initial popularity (pledged money) was low, but the pledged money gradually increased until the 16th time bucket with sharp increments between 12th and 14th time buckets. A cluster C1 (less interesting cluster) had a similar pattern with C4, but overall increments were much lower than C4.

- A cluster C5 had also an interesting pattern which was gradually going up during the first half duration and going down during the other half duration.

Next, we analyzed how many projects belonged to each cluster, and estimated average project goal and pledged money of projects in each cluster. Table 4.1 shows the number of projects, and corresponding average project goal and average pledged money. Two largest clusters were C2 and C1 consisting of 28,209 (93%) and 1,563 (5%) projects, respectively. These clusters had the lowest goal and achieved the lowest pledged money compared with the other three clusters. C3 had the highest goal and got the highest pledged money. C4 and C5 had next highest goal and got next highest pledged money. Overall, Each of the top 2% successful projects (including C3, C4, and C5) on average received more than 200K pledged money. It means that there were a lot of successful projects with a low goal and low pledged money, while there existed a small portion of projects (2%) with a high goal and high pledged money, resulting in unequal distribution of pledged money across successful projects in a crowdfunding platform, Kickstarter.

[1] The Coolest Cooler project received $13,285,226, and the Pono Music project received $6,225,354.

Table 4.1: Number of projects, average project goal and average pledged money in each cluster.

Cluster	|projects|	Avg. goal	Avg. pledged money
C1	1,563	$41,542	$95,429
C2	28,209	$6,334	$9,306
C3	97	$273,222	$1,487,672
C4	186	$98,253	$227,078
C5	278	$79,354	$284,761

Table 4.2: Average percent of duration reaching a goal in each cluster.

Cluster	Avg. goal	Avg. percent of duration reaching a goal
C1	$41,542	55%
C2	$6,334	66%
C3	$273,222	17%
C4	$98,253	58%
C5	$79,354	26%

Up-coming questions are "When did projects in each cluster reach their goal? Did they reach in almost similar time (e.g., the first 30% duration)?". To answer these question, we analyzed accumulated daily pledged money to see when they reached the goal. Table 4.2 presents the analytical results. All the successful projects reached their goal before 67% duration. Projects in cluster C3 (with the highest goal and pledged fund) reached their goal very fast, only in 17% duration. Projects in C5 reached their goal faster than projects in C4, but total pledged money was less than C4 at the end of the fund-raising campaigns. Interestingly, projects in C1, which had similar (but less popular) temporal pattern with C4 in Figure 4.2, reached their goal in similar time (55%) even though their goal was lower than C4. C2 with the lowest goal took the longest duration to reach the goal.

Next, we further analyzed the five clusters to understand how other properties were

Table 4.3: Average property values in Successful clusters.

Cluster	Average						
	Pl. Money	|Images|	|Videos|	|FAQs|	|Rewards|	|Updates|	|Comments|
C1	85,429	18.36	2.03	3.38	15.51	19.94	405.70
C2	9,306	6.59	1.28	0.72	10.07	9.14	26.89
C3	1,487,672	34.44	2.51	12.71	18.20	41.80	16,712.34
C4	227,077	23.74	2.52	5.50	18.89	27.28	1509.78
C5	284,761	22.24	2.20	7.66	14.57	23.94	1233.47

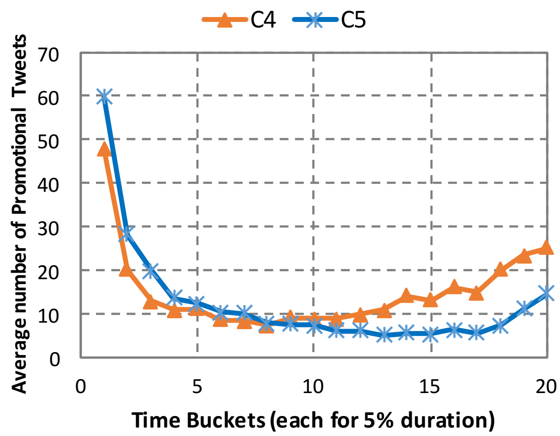


Fig. 4.3: Average number of promotional tweets posted during each time bucket in C4 and C5.

associated with pledged money across the five clusters. In particular, we focused on properties such as the number of updates, the number of images, the number of videos, the number of FAQs, the number of rewards, the number of updates and number of comments. Table 4.3 shows the average value of the properties in each cluster. We clearly observed that projects in C3 had the largest values in all the properties except the number of videos (still almost similar with the largest value in C4). Project creators in C3 spent more time to create their project descriptions by adding more images, videos and reward types. During a fund-raising period, they actively added more updates, FAQs and received more comments from backers. Mostly, these phenomena applied to the other clusters.

Finally, we focused on C4 and C5 which had interesting evolutional patterns as shown in Figure 4.2. Specifically, projects in C4 were initially not popular, but later became popular with a sharp increment in terms of relative pledged money in each time bucket, while projects in C5 were initially popular and then became less popular or relative pledged money in each time bucket decreased. To understand the phenomenon, we investigated how external promotional activities in C4 and C5 were different.

To conduct this study, first, we collected promotion-related tweets for each project in C4 and C5 from Twitter by searching each Kickstarter project URL. These tweets were posted by project creators, their friends, and backers. Then, we computed the average number of promotion tweets during each time bucket in each cluster. Figure 4.3 shows how the number of promotion tweets was changed over time. Interestingly, in the first 8 time buckets, the number of promotion tweets in C5 were higher than the number of promotion tweets in C4. Since then, the situation was reversed – there were more promotion tweets in C4 than C5. Interestingly, the temporal promotional activities were similar with the evolutional patterns of pledged money in C4 and C5 shown in Figure 4.2. Note that it took time for these promotional activities to take effect in terms of relative pledged money in each time bucket. Based on this study, we conclude that promotional activities on social media played an important role in increasing relative pledged money over time.

CHAPTER 5

PROJECT CREATOR'S REACTIONS AFTER PROJECTS FAILED

In this section, we analyze how project creators behaved after their projects failed. Did they give up and no longer create projects? Or did they continue to create projects? If they continued creating projects with the same idea of the failed projects, what changes did they make in order to make the projects successful?

First of all, we analyzed how many projects each user created in Kickstarter as shown in Table 5.1. 89.74% (118,718) users created only 1 project while 7.97% users created 2 projects and 2.29% users created at least 3 projects. Among the 89.74% creators, who created only 1 project, 44.15% project creators successfully reached project goals (i.e., fund-raising goals) while 55.85% project creators failed in reaching project goals. It may mean that the 55.85% (66,304) project creators among the one-time project creators gave up their project idea, and no longer created new projects.

A follow-up question is "when a project failed, what properties of the project did project creators change to make the project successful?" Did they lower project goal? or Did they add more reward types? or Did they add more detailed information into the project description? Before answering these questions, we assume that once a certain project is successful, the project creator will no longer improve or relaunch it. But if a project failed, the project creator may (i) want to improve and relaunch it, (ii) create a project with a completely new idea, or (iii) no longer create any other project. In this study, we focus on the first (i) case because we aim to understand what properties of the previously failed project the project creators changed to make it (of the same idea with the previous project) successful.

A challenge in the study was to extract two consecutive projects based on the *same project idea* in chronical order. We assumed that if two consecutive projects created by the same creator were based on the same idea, their project descriptions should be similar. Based on this assumption, we examined 22,320 projects created by 9,166 distinct creators, each of whom created at least 2 projects and had at least one failed project. Then we built

Table 5.1: Distribution of projects by creators.

# created project	# creators	Percentage (%)
1	118,718	89.74
2	10,546	7.97
3	1,959	1.48
4	546	0.41
5	235	0.18
> 5	282	0.21

Vector Space Model for 22,320 projects so that each project was represented by a TF-IDF based vector [26]. We extracted each pair of two consecutive projects created by the same user from the 22,320 projects and measured the cosine (description) similarity of the pair.

Specifically, given two projects P_i and P_j represented by two vectors V_i and V_j respectively, cosine (description) similarity was calculated as follows:

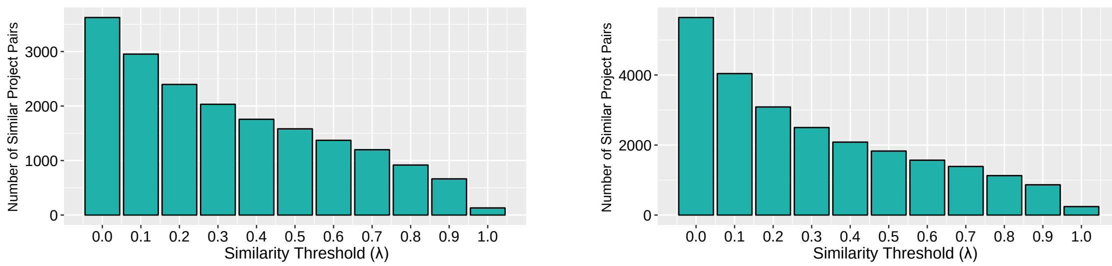
$$sim(P_i, P_j) = cos(V_i, V_j) = \frac{\sum_{k=1}^{|D|} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{|D|} v_{ik}^2} \sqrt{\sum_{k=1}^{|D|} v_{jk}^2}}$$

where, $|D|$ is the total number of unique terms in Vector Space Model, v_{ik} and v_{jk} are TF-IDF values at k^{th} dimension of V_i and V_j, respectively.

If a pair's cosine similarity was equal to or greater than a threshold λ, we would consider the pair as similar projects based on the same project idea.

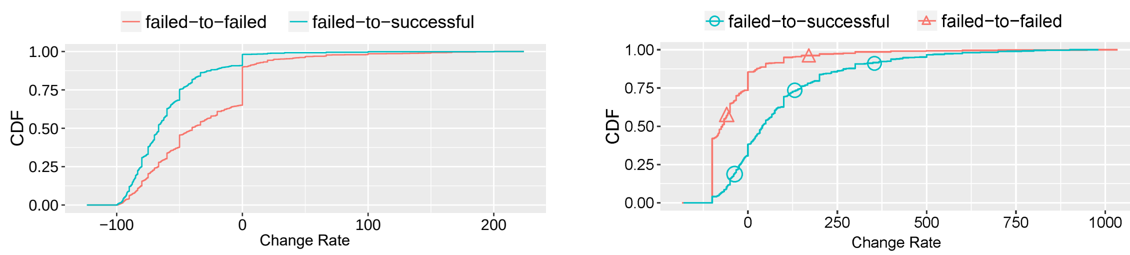
An up-coming question is what would be a good λ? To answer this question, first we plotted Figure 5.1 which shows how the number of pairs of failed-to-failed projects and the number of pairs of failed-to-successful projects were changed as we changed λ from 0 to 1 by increasing 0.1. The number of similar project pairs had decreased as we increased λ. Interestingly, we observed that there were 131 pairs and 242 pairs of projects without changing any word in their project descriptions (i.e., similar score = 1) in Figure 5.1a and Figure 5.1b, respectively. It means some project creators did not change project description of the latter project compared with the former project, but it was successful in 131 cases.

Then, we manually analyzed sample pairs to see what threshold would be the most appropriate to find similar project pairs. Based on the manual investigation, we decided λ as 0.8. With the threshold (λ=0.8), we found 918 failed-to-successful project pairs called *group*



(a) Number of failed-to-successful project pairs. (b) Number of failed-to-failed project pairs.

Fig. 5.1: Number of similar project pairs in failed-to-successful case and failed-to-failed case.



(a) Goal (b) Number of updates

Fig. 5.2: CDFs of change rates of goal and number of updates in similar project pairs.

I and 1,127 failed-to-failed project pairs called *group II*. By comparing projects in each pair in the two groups, we noticed that overall project creators changed 13 properties: duration, goal, number of images, number of videos, number of faqs, number of updates, number of rewards, number of sentences in reward description, smog grade of reward, number of sentences in project description, smog grade of project description, number of sentences in project creator's biography, and smog grade of project creator's biography. We measured how much each property was changed by $\frac{(P_{ik}-P_{jk})*100}{P_{ik}}$ where P_{ik} is the former project's kth property value and P_{jk} is the latter property's kth property value.

Table 5.2 shows the average change rate of failed-to-successful project pairs and failed-to-failed project pairs. A positive change rate means that project creators increased the property value of the latter project compared with the former project. To measure which property had the significant difference, we computed one-tailed p-value of two-sample t-test for difference between the means of the two groups. In particular, the mean of project goal's change rate in *group I* was -59.62%, which was approximately four times decrement

Table 5.2: Average change rate of 13 properties in failed-to-successful project pairs and failed-to-failed project pairs. $***$, $**$, $*$ and ns indicate $p < 10^{-13}$, $p < 10^{-4}$, $p < 0.05$ and *not significant*, respectively.

Property	Avg. change rate of failed-to-successful pairs *Group I*	Avg. change rate of failed-to-failed pairs *Group II*	p-value
Duration	-6.15%	+23.03%	**
Goal	**-59.62%**	**-16.39%**	***
#images	+14.25%	+1.91%	*
#video	+6.40%	-3.22%	**
#faq	-34.69%	-47.47%	ns
#reward	-0.26%	+2.36%	ns
#updates	**+118.00%**	**-38.41%**	***
smog_reward	+1.70%	+2.78%	ns
#reward_sentence	+22.24%	+13.73%	ns
#main_sentence	-0.40%	-0.27%	ns
smog_main	+7.26%	+5.17%	ns
#bio_sentence	0%	0%	ns
smog_bio	0%	0%	ns

compared to *group II* which had -16.39% change rate. In other words, project creators in *group I* lowered project goal much more than project creators in *group II*. The mean of the change rate of the number of updates in *group I* was +118% while project creators in *group II* made -38.41% change. It indicates that project creators in *group I* increased the number of updates significantly while project creators in *group II* decreased the number of updates. Interestingly, decreasing a project duration was helpful to make projects successful. Overall, reducing the duration and goal as well as posting more images, videos and updates are a smart way to make previously failed projects successful.

Since the number of updates and project goal were the most significant properties, we further analyzed CDFs of change rates of the two properties – project goal and a number of updates – in the two groups as shown in Figure 5.2. 88% project creators in *group I* lowered project goal while 63% project creators in *group II* lowered project goal. About 62% project creators in *group I* increased posting the number of updates while only 15% project creators in *group II* increased posting the number of updates.

CHAPTER 6

ANALYZING POST CAMPAIGN ACTIVITIES ON KICKSTARTER

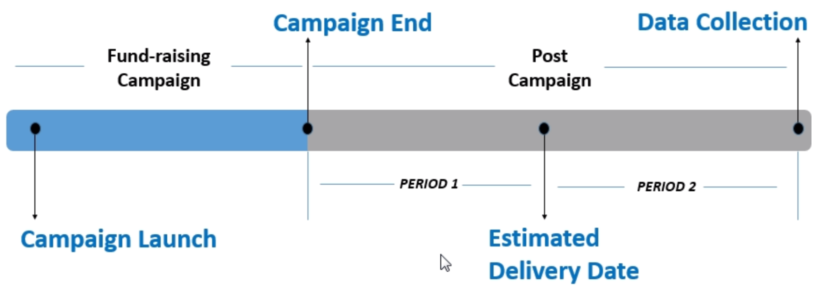


Fig. 6.1: Timeline of a project in Kickstarter.

In the previous chapters we analyzed the characteristics of successful projects by clustering them and also observed how project creators reacted when their projects failed. In this chapter we mainly focus on analyzing the post campaign activities. Before we analyze the post campaign activities, first, we need to understand what happens after the fund-raising campaign. After the project is successful, the creators are supposed to deliver all the promised rewards to the backers within the estimated delivery date set by the creator during the project launch. Authors, in a case study of donor retention in crowd-funding platforms ([11]), examined that if creators send all rewards to the backers on time, they will set a trusted image in backers thought and there would be a higher probability that backers may back for his next projects in future. In contrary, if the creator delays reward delivery, the backers feel disappointed and may no longer back for his next projects. Firstly, from Figure 6.1 we can distinguish a post campaign from fund-raising campaign. The duration of fund-raising campaign ranges from the project launch date to the project end date whereas the campaign after the project end date is called as post-campaign.

Hence, in the below sections we observe whether creators delivered all the promised rewards within the estimated delivery date and further distinguish those projects as on-time delivered projects if the creator delivers all the rewards within the estimated delivery date and if the creator delivered all the promised rewards after the estimated delivery date, we call them as late-delivered projects and also we analyzed the characteristics of the backers and creators in post-campaign.

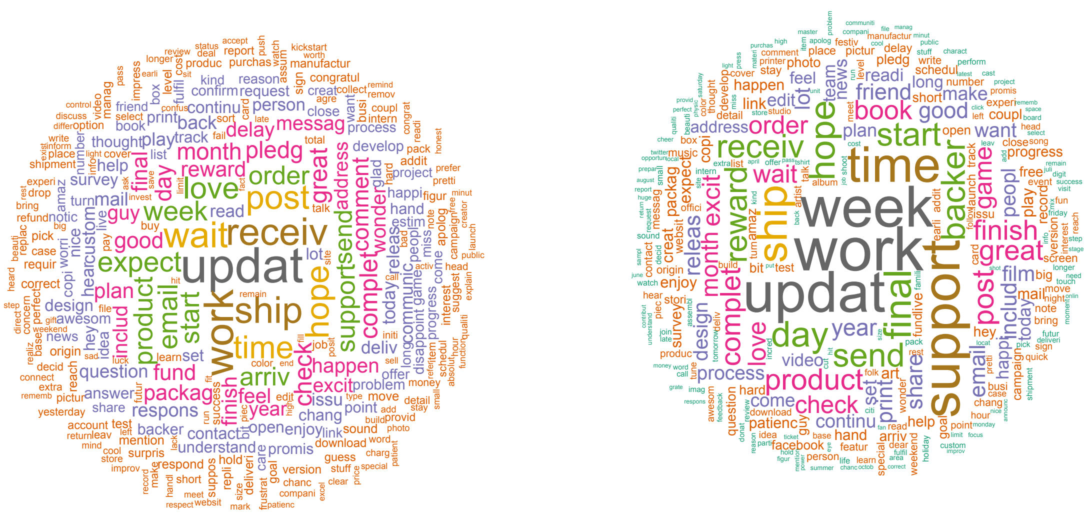
6.1 Preprocessing the data

As we are interested in analyzing post campaign activity, we used 39,972 project pages, a subset of 151,608 project pages as mentioned in Chapter 3, which were successful between 2009(the Kickstarter launched year) and October 2013. The Kickstarter [27], in its terms of use, mentioned that the date listed on each reward is the creator‘s estimate of when they will provide the reward. Hence for observing on-time and late delivered projects, the golden truth is the estimated delivery date which is set by the creator. Thus, we considered the projects that have estimated delivery dates defined. The dataset is further filtered to 29,499 projects by removing the noisy projects that have goal less than 1,000$ and greater than 1,000,000$.

6.2 Ground truth

Since there is no publicly available ground-truth data of whether the rewards are delivered to backers on time, we build the ground truth by labeling the projects as ‘on-time delivered’ and ‘late time delivered’ based on the comments, updates and estimated delivery date (EDD.) The main reason for considering comments and updates as main decision makers is because, thorough observation of comments and updates shows that backers often discuss about the products, rewards, updates, shipping status, quality of product and creators discuss about the production of product, shipping dates and survey. Figure 6.2a and Figure 6.2b shows the word clouds of comments and updates in Kickstarter. The most popular terms used are about shipment (e.g., receive, ship, package, customs and shipment), update request or posting updates (updates, post), compliments (great, good, awesome), and asking refund. Some of the sample comments are shown in Table 6.1.

We have considered the EDD for a project as the maximum of all the reward’s EDD for the project. Given a set of comments, updates and max. EDD for the project we define criteria and then label them based on the defined criteria. We observe the updates and



(a) Word cloud for comments in Kickstarter.

(b) Word cloud for updates in Kickstarter.

Fig. 6.2: Word clouds of comments and updates in Kickstarter projects.

comments made by creator(only) to decide the delivery status.

A project is labeled as On-time delivered if it satisfies at least one of the following conditions:

- If it has at least one update before EDD stating that rewards are sent or
- If it has comments made by creator saying that rewards are sent.

A project is labeled as Late delivered if it satisfies at least one of the following conditions:

- If it has at least one update after EDD stating that rewards shipping is delayed or
- If it has comments made by creator excusing for the delay in the rewards. (He may include a new estimated delivery.)

We don't consider the projects for labeling if they do not follow the above rules and we have excluded the projects that neither have updates nor comments from creator.

Two human labelers independently classified each project to either on-time delivery or late delivery based on the above mentioned criteria and achieved 98% agreement. Since labeling all the 29,499 projects takes longer time, we randomly selected 2,949 projects.

Table 6.1: Sample of Backer and creator comments

Comments	Topic
Backer Comments	
I got mine it is really beautiful. Thank you so much	Compliments
Congratulations on successful funding Im so glad to be a part of this excellent project and cannot wait to see the finished film	Compliments
Survey completed now watching for the postman	Survey
Yes an update would be greatly appreciated	Requesting Updates
How come there were no recent updates	Requesting Updates
How do I go about getting a refund from amazon or is it too late reading the comments and researching the company more? It seems to me this project is starting to smell fraudulent. I'm not very familiar with this process since its my first and now last backed project Kickstarter is a wonderful platform but after this experience. I'm not coming back.	Refund
I am also waiting for the creator to come on and explain the difference between anodized and cerakote	Product
Creator Comments	
Awesome glad you really liked it	Compliments
I'm going to post a new update with links for all the packs. Sorry that some of you haven't got the emails in a timely manner I'll check my dB to make sure I didn't mess up your address	Update
Michael we have personally addressed your comment in a message. The next shipment of Nikons will arrive in April sit tight.	Shipping
All units will be shipped usps	Shipping
Thanks for hanging in with me through all of this guys. It has been hell but we made it. I wont stretch your patience again	Thank you note
My apologies it's not failed I am going to make an update	Apologies for being late
Surveys will be going out in early august. The rewards will ship around the release date October.	Survey

After eliminating the projects that does not satisfy the above defined criteria we have 2,198 projects.

6.3 Topic Modelling

Studying the characteristics of comments in on-line social networking sites, forums, discussion boards, micro blogging sites and commercial sites becomes important for a number of tasks such as topics extraction, sentimental analysis and breaking news detection. Fortunately, Kickstarter also provides a platform for communication among creator and backers for the project.

This section answers the following question "What are the topics that interested the people most or talked about after the fund-raising campaign?" Answering this might help us in knowing the key issues of the crowd-funding platform after the campaign is successful. To answer this question we have extracted the comments left by both creators and backers after campaign end date. As mentioned in the previous section we have extracted the maximum estimated delivery date of the rewards for each project, thereby we divided the time period between the campaign end date and till date into two different time-lines. Hence we can analyze the topics in crowd-funding over the time. These time-lines are shown in the Figure 6.1. The Period 1 is the duration between Campaign end date and Estimated Delivery Date and the period after the Estimated Delivery Date is described as Period 2.

6.3.1 Topic models

We have used Latent Dirichlet Allocation model to discover the topics in comments collection. Latent Dirichlet Allocation is an unsupervised machine learning technique to identify latent topics in document collections. It uses 'bag of words' approach, which treats each document as a vector of word counts. Each document is represented as a probability distribution over a number of words.

LDA is defined more formally as, each document in the collection is associated with a multinomial distribution over T topics, which is denoted as θ. Each topic is associated with a multinomial distribution over words, denoted as ϕ. Both θ and ϕ have Dirichlet prior with hyper parameters α and β respectively. For each document d, a topic Z is sampled from the multinomial distribution θ associated with the document and a word from the multinomial distribution ϕ associated with the topic Z is sampled consequently. This process is repeated N_d times where N_d is the total number of words in the document d.

6.3.2 Modeling the comments

As mentioned above we have extracted comments made by backers and creators in two different time-lines and the distribution of these comments in two different periods is shown in Table 6.2.

Prior to extracting the topics we implement the following preprocessing steps: 1) Remove the links from comments, 2) Remove any words starting with '@' character (used to

Table 6.2: The distribution of the comments among creators and backers in two time-lines.

	Backer's Comments	**Creator's Comments**
Period 1	49115	4513
Period 2	43160	5283

Table 6.3: Top 5 interesting topics among creator and backer's comments.

Topic	**Keywords**
	Backer-Period 1
Appreciating the idea	good wait love time awesome game great work happy guys cool hope play kickstarter project nice idea hey lol stuff
Unhappy backers asking Refund	kickstarter project product money people backers projects backed company funding months refund understand feel issues campaign unhappy products business delays
Backer address survey	survey email received address shipping kickstarter send receive contact question pledge check info order backer wondering reward message response rewards
Expressing happiness on reward delivery	today received arrived great package box yesterday wait mail love copy quality work awesome week live book days waiting day
Product production progress	update updates shipping delivery time news wait month hope ship week backers waiting project days progress production early weeks guys
	Backer-Period 2
Acknowledging shipping information	tracking number received shipping waiting customs shipped package receive days pay today backer send post usps list shipment email delivered
Requesting updates	update updates backers project shipping time months kickstarter delivery news hope month waiting weeks week progress shipped ship expect communication
Backers requesting refunds	refund kickstarter project money backers creator card projects creators reward refunds credit terms rewards request backer fulfill legal amazon action
Discussion about Products	work working device works screen button app fine phone time software android problem wifi mode running power issues usb connect
Fraudulent activity	money kickstarter refund people scam fraud project year lies care comment company post deliver backers product promised paid crap bullshit
	Creator-Period 1
Feedback and support	day happy glad comments questions great hear love feedback support problem campaign wait feel free kind answer account future kickstarter
Updates about production	update time working week backers post work good hey updates guys production days today lot ready coming bit people posted
Survey on person specific rewards	survey size color manufacturer choice colors confirm green choose june material samples review sizes black white submitted options selected chart
Shipping information	shipping send ship week shipped rewards survey delay product issues info address days case box production issue delivery schedule arrive
Payment problems	paypal problem answer account point main option understand question money play pay mind cost payment including real asked thought include
	Creator-Period 2
Feedback and support	support questions great comment backers issue comments work future happy shipping feedback contact hope issues feel share question free product
Delayed creators requesting patience from backers	update patience hey post couple coming kickstarter long posted rewards delay process guys waiting longer updates news support weeks final
Acknowledging shipping information	message received orders international tracking package customs kickstarter packages happy private arrive shipped batch receiving response delayed shipment drop inbox
Describing the difficulties faced in implementation	covered refunded industrial properly design forwarder partner repair investigate reference models victor jun wear chen matthias challenges pledges manual stefan
Money back for damaged products	trout behavior brett stuff damages books project send professional damaged legal money texas iowa law cool midnight harassment toys caused

refer the users. e.g., @belkey,) 3) Remove all non-Latin characters from the comments and covert all words to lower case and 4) Remove all stop words.

After preprocessing the comments we run the LDA model on each set of collection of comments to find the topics over time. We used perplexity as a metric to choose the best number of topics. The calculated perplexities for every collection is shown in Figure 6.3.

We ran the LDA model on every collection of comments with number of topics ranging from 1 to 40 and got the optimal number of topics as 36 for Backer-Period 1 and 39, 36 and 28 for Backer-Period 2, Creator-Period 1 and Creator-Period 2 respectively. The five

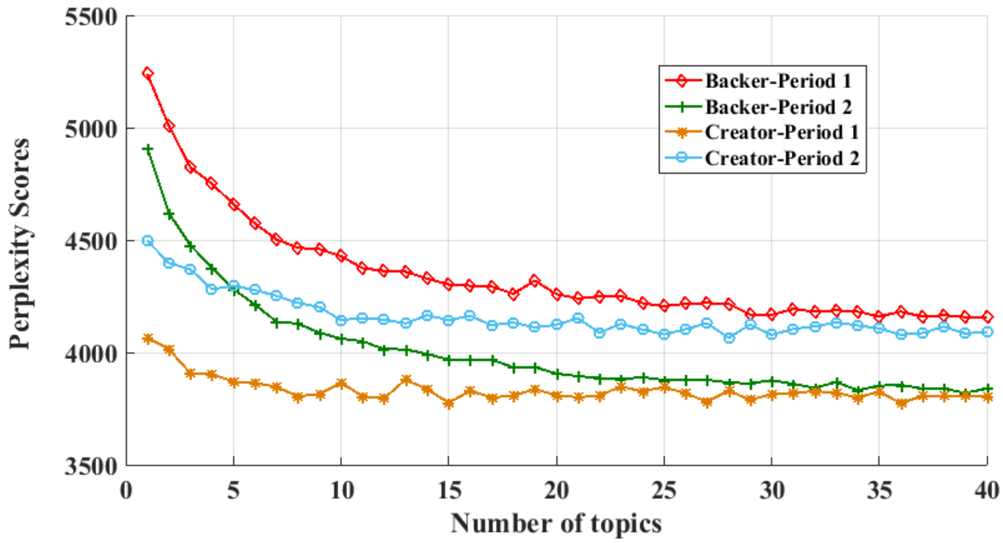


Fig. 6.3: Perplexities scores for every collection of comments.

top most interesting topics from every collection is shown in Table 6.3.

6.3.3 Analysis of the most discussed topics over time

Labeling the topics is an unsupervised technique as we don't have gold-standard topics. Hence, we carefully look through the keywords and read the comments that has the particular keyword and then label them to appropriate topics.

Backer Period 1

From Table 6.3 we see that backers in period 1 are mainly discussing about appreciating the creator's idea, discussing about email surveys regarding backer's addresses for rewards delivery, expressing happiness about rewards delivery while unhappy backers who didn't receive any updates and rewards are asking for refunds and discussing the progress of product's production.

Backer Period 2

In period 2 backers mainly talk about fraudulent projects which does give any updates to backers even after continuously contacting creators. They also acknowledge creators and other backers upon their shipment delivery and often express happiness and share the characteristics of the rewards. We also observed that in period 2 backers are requesting refunds if they are unsatisfied with the creator's activeness or if they don't receive any rewards after estimated delivery dates. Backers often posted about the quality of products and asked questions about working of products.

Creator period 1

Table 6.3 shows the interesting topics from creator's comments, in period 1, and corresponding keywords obtained from topic modeling. We observed that creators ask feedback and support from backers and requests to ask any questions about the product. Creators also post updates about the production process and share about their working schedule. They also conduct surveys to know the specific characteristics (e.g., color of shirt, size of shirt) from backers as some rewards might be customized. Creators inform backers about shipping schedule and short delays. We can see in the Table 6.3, that there is a topic called payment problems and we can say that here creators discussed and answered questions about the payment problems in Kickstarter.

Creator period 2

In period 2 creators mainly talk about feedback and support from backers as in period 1 and also requesting patience from backers say they could not deliver the rewards on time. The rewards might have been sent in batches and they provide the shipping schedules for all the batches. If the rewards are not delivered on time they describe the challenges faced by them in production and propose new delivery dates. Another interesting topic is about refunds for damaged products.

From the above analysis, we can say that backers who are satisfied with the rewards are expressing their happiness by thanking creators while backers who are unhappy with creators request updates or refunds and also they talk about fraudulent projects where they don't receive updates or comments from creators. On the other hand, creators continuously update backers about progress and production of the project, the shipping schedule and request support and patience if they cannot finish and deliver projects on-time.

Table 6.4: Distribution of projects among categories

	Category	**Projects**
1	Art	222
2	Comics	85
3	Crafts	12
4	Dance	66
5	Design	130
6	Fashion	76
7	Film	581
8	Food	113
9	Games	155
10	Journalism	9
11	Music	740
12	Photography	55
13	Publishing	224
14	Technology	60
15	Theater	163

6.4 Analysis of the projects in post-campaign

Table 6.4 shows the distribution of 2,198 projects among 15 categories. We observed top 5 categories with high number of projects in music, film, art, publishing and theater. Journalism and crafts category are two categories with lowest number of projects. In this section, we analyze projects in post campaign to answer following research questions: Do we observe different proportions of late-delivered projects in different categories? Does setting longer delivery duration avoid lately delivering all rewards?

To answer the first question "Do we observe different proportions of late-delivery projects in different categories?", we plot the distribution of on-time delivered and late delivered projects in Figure 6.4. We observed that games, comics, technology and design categories have higher proportion of late-delivery projects with 78%, 65%, 64% and 64% respectively. In contrast, only 7% and 10% projects belong to dance and theater categories are late-delivery projects. The rest of categories have range of proportion of late-delivery projects from 32% to 55%. We speculate that promised rewards of projects belong to dance or theater category are just live performance, showcase or teaching class (e.g. dancing class) which served all backers in once. Thus, it is easier to deliver the promised rewards on time. However, creators in projects belong to other categories must produce real products (e.g. T-shirts, photo album, book or e-books in fashion, photography, comics category respectively) and send to backers one by one. This process takes longer time than what they estimated and thus caused higher proportion of late-delivery projects in such categories.

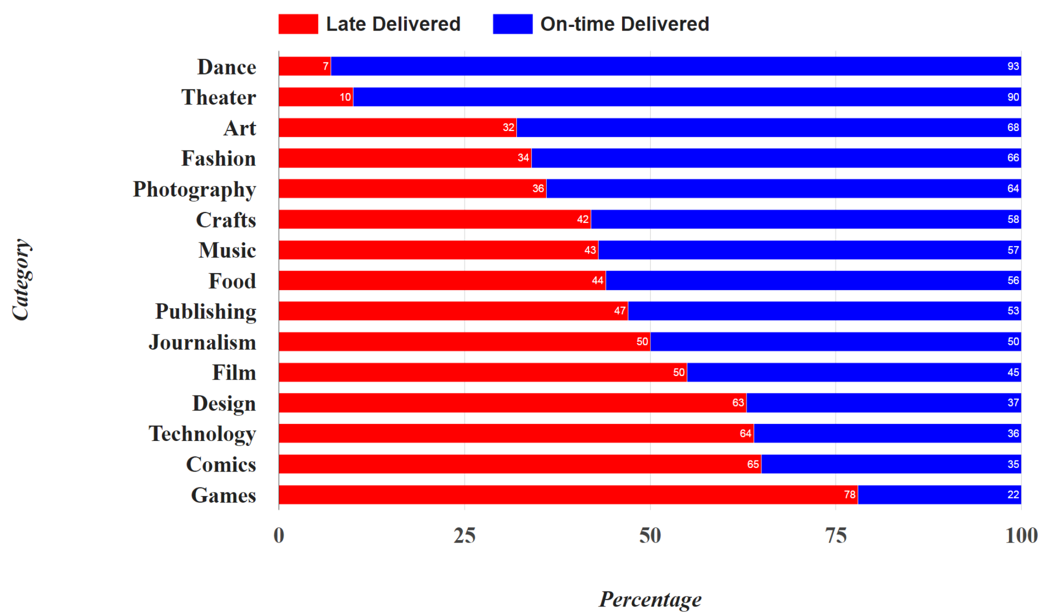


Fig. 6.4: Distribution of on-time delivered and late-delivered projects over 15 categories

To understand where setting longer estimated delivery duration can reduce rate of late delivery, we plot the distribution of on-time delivered and late delivered projects with different estimated delivery duration ranges in Figure 6.5. Most projects set the estimated delivery duration from 2 - 6 months. Interestingly, we observed that for the projects whose delivery duration ranges from 1 - 9 months, the percentage of late delivered projects is dominated by on-time delivered projects. In contrast, the percentage of late-delivered projects is dominated by on-time delivered projects in projects with delivery duration range greater than 10 months. Hence, this suggests that setting longer estimated delivery dates helps

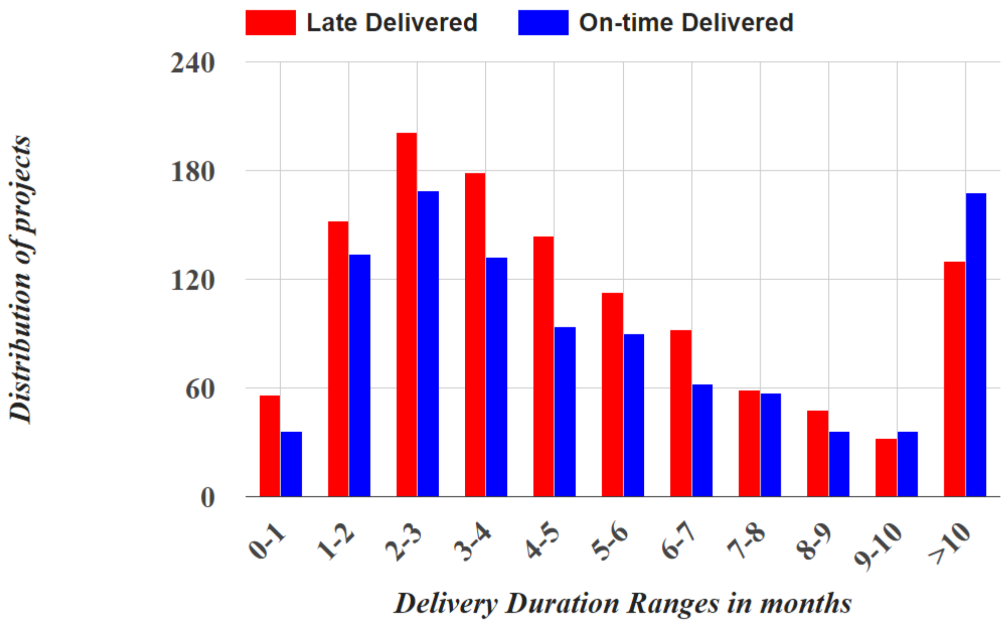


Fig. 6.5: Distribution of on-time delivery and late-delivery projects with different estimated delivery duration ranges

creators deliver the rewards on-time.

6.5 Analysis of the backers and creators in post-campaign

In this section we analyze the distinguishing characteristics of creators and backers in on-time and late delivered projects. Here, we answered following research questions: "How active are the backers in On-time and Late delivered projects?" and "How fast did the creators comment back to backer's comments?"

To answer the question, "How active are the backers in On-time and Late delivered projects?" we have calculated the ratio of number of backers who commented on the project's comment section to the total number of backers who backed the project for on-time delivered projects and late delivered projects respectively. We compared the mentioned ratio in fund-raising and post campaigns to observe the change in activeness of backers. Figure 6.6 shows the CDF of ratio of number of backers who commented on the project's comment section to the total number of backers who backed the project in both fund-raising and post campaigns. Here, we can observe that: (1) the backers are more active in fund raising

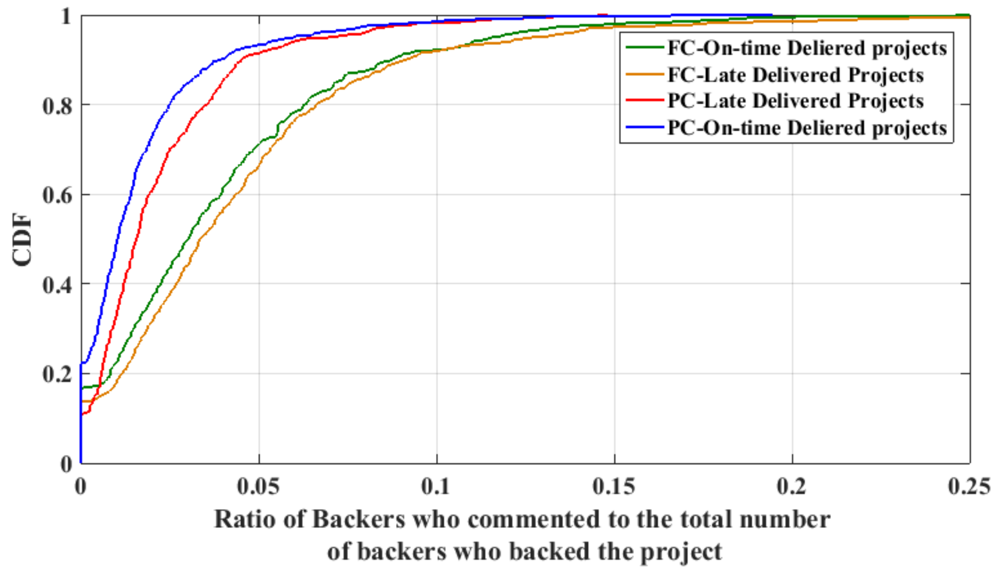


Fig. 6.6: CDF of Ratio of backers who commented and the total backers

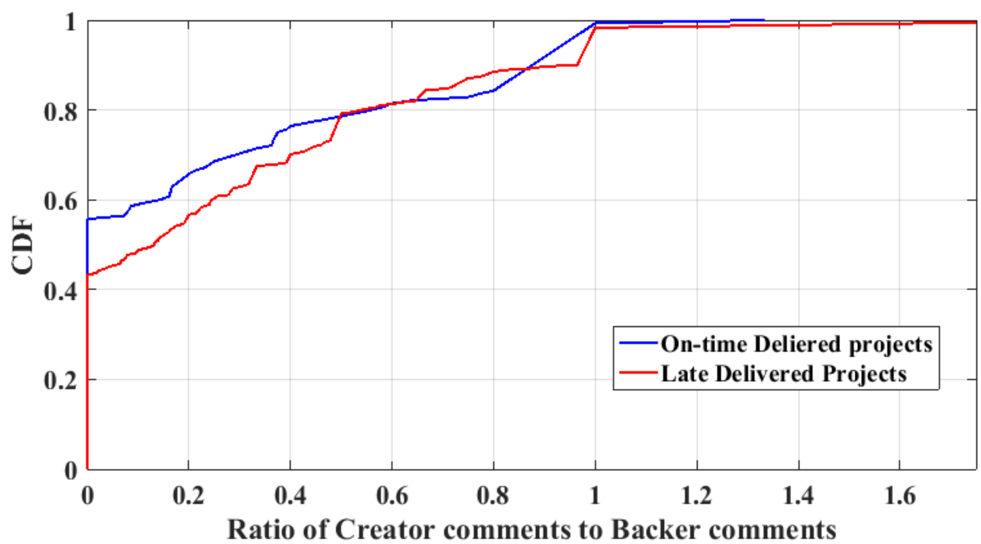


Fig. 6.7: CDF of Ratio of Creator to Backer comments in a post campaign

campaign compared to post campaign, (2) in fund-raising and post campaigns late delivered projects have more comments compared to on-time delivered projects. In particular, 90% of the projects in post campaign, have at most 3.9% and 5% of backers who commented, in on-time delivered projects and late delivered projects respectively. And, 90% of the projects in fund-raising campaign, have at most 8% and 9% of backers who commented, in on-time

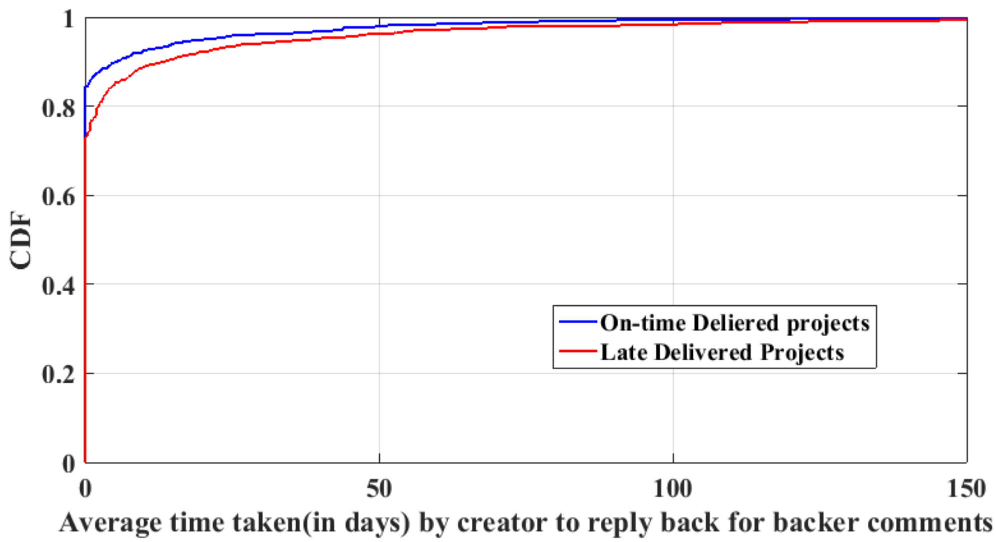


Fig. 6.8: CDF of Average time taken by a creator to comment back to Backers

delivered projects and late delivered projects respectively. This explains that backers are more active in Late delivered projects and from the topics extracted in topic modeling we can speculate that they might be discussing about topics like requesting updates, requesting refunds as the projects are not delivered on-time and fraudulent activities.

Next, we analyzed the ratio of Creator comments to the Backer comments in the post campaign for on-time and late delivered projects. From Figure6.7, the ratio is higher in late delivered projects than on-time delivered projects which suggests that comparatively creators in late delivered projects commented back to backer comments more than the creators in on-time delivered projects. In 70% of the projects, creators in late delivered projects commented once for every 2 comments from backers where as in on-time delivered projects creator comments 3 times for every 10 backer comments. From Figure 6.6 we have seen that backers from late delivered projects are more active and since, there are more backer's comments, creators have to comment back more.

Another interesting question is "How fast did the creators comment back to backer's comments?" To answer this question we have calculated the average time taken by the creator to comment back to backer's comments for every project in on-time and late delivered projects. Figure 6.8 clearly shows that creators in late delivered projects took longer time to reply back to backer's comments than the creators in on-time delivered projects. Surprisingly, in 90% of the projects in on-time delivered projects, creators an average of 8 days to reply back to backer's comments where as in late-delivered projects, creators took an average of 26 days to comment back. The interesting thing is that, even though, creator to backer comments ratio is higher in late delivered projects, creators here took very long time to reply back. This explains that backers in late delivered projects are more active and comments very high in number compared to on-time delivered projects.

CHAPTER 7
CONCLUSION AND FUTURE WORK

In this research, we have clustered successful projects based on a time series of relative pledged money, and found 5 clusters. Out of the 5 clusters, we found three interesting clusters: (i) projects in a cluster were the most popular, receiving the highest relative pledged money over time; (ii) relative pledged money of projects in a cluster went up and went down; and (iii) relative pledged money of projects in a cluster had low relative pledged money initially, but went up with a sharp increment. Next, we analyzed what reactions project creators made when their projects failed. By identifying similar project pairs, we compared what properties project creators changed in order to make their failed projects successful in the next try. Our t-test revealed that project creators who lowered their project goal by -59.62% and increased posting the number of updated by +118% on average made the projects successful.

We also explained the importance of creators in post campaign and build the ground truth to differentiate the projects as on-time delivered and late delivered projects by labeling them based on defined criteria. We performed topic modeling of comments to understand the topics that backers and creators talked about after the fund raising campaign. We found that backers mostly ask updates and progress of the project and if they don't get their within the anticipated time they either ask for updates about new delivery date or request refunds. And, creators update backers about progress and production of the project, the shipping schedule and request support and patience if they cannot finish and deliver projects on-time. Next, we also analyzed the distinguishing characteristics of on-time delivered and late delivered projects and observed the activeness of backers and creators in post campaign.

There are few future research explorations we may explore using these features, we can build prediction models to predict whether project creators deliver rewards to backers on time and also suggest the feasible estimated delivery dates for the rewards delivery.

REFERENCES

[1] Lu, C.-T., Xie, S., Kong, X., and Yu, P. S., "Inferring the impacts of social media on crowdfunding," *Proceedings of the 7th ACM international conference on Web search and data mining*, ACM, 2014, pp. 573–582.

[2] Chung, J. and Lee, K., "A Long-Term Study of a Crowdfunding Platform: Predicting Project Success and Fundraising Amount," *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, ACM, 2015, pp. 211–220.

[3] CrowdExpert, "CROWDFUNDING INDUSTRY STATISTICS 2015 2016," http://crowdexpert.com/crowdfunding-industry-statistics/, January 2016.

[4] Belleflamme, P., Lambert, T., and Schwienbacher, A., "Crowdfunding: Tapping the Right Crowd," *SSRN Electronic Journal*, 2012.

[5] Gerber, E. M. and Hui, J., "Crowdfunding: Motivations and Deterrents for Participation," *ACM Trans. Comput.-Hum. Interact.*, Vol. 20, No. 6, Dec. 2013.

[6] Gerber, E. M., Hui, J. S., and Kuo, P.-Y., "Crowdfunding: Why People Are Motivated to Post and Fund Projects on Crowdfunding Platforms," *CSCW*, 2012.

[7] Hui, J. S., Greenberg, M. D., and Gerber, E. M., "Understanding the role of community in crowdfunding work," *CSCW*, 2014.

[8] Kuppuswamy, V. and Bayus, B. L., "Crowdfunding creative ideas: The dynamics of project backers in Kickstarter," *UNC Kenan-Flagler Research Paper*, , No. 2013-15, 2015.

[9] Mollick, E., "The dynamics of crowdfunding: An exploratory study," *Journal of Business Venturing*, Vol. 29, No. 1, 2014.

[10] Joenssen, D. W. and Müllerleile, T., "Limitless Crowdfunding? The Effect of Scarcity Management," *Crowdfunding in Europe*, Springer, 2016, pp. 193–199.

[11] Althoff, T. and Leskovec, J., "Donor Retention in Online Crowdfunding Communities: A Case Study of DonorsChoose. org," *WWW*, 2015.

[12] Etter, V., Grossglauser, M., and Thiran, P., "Launch hard or go home!: predicting the success of kickstarter campaigns," *Proceedings of the first ACM conference on Online social networks*, ACM, 2013, pp. 177–182.

[13] Greenberg, M. D., Pardo, B., Hariharan, K., and Gerber, E., "Crowdfunding support tools: predicting success & failure," *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2013, pp. 1815–1820.

[14] Zimmerman, J., Forlizzi, J., and Evenson, S., "Research through design as a method for interaction design research in HCI," *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2007, pp. 493–502.

[15] Lu, C.-T., Shuai, H.-H., and Yu, P. S., "Identifying your customers in social networks," *CIKM*, 2014.

[16] An, J., Quercia, D., and Crowcroft, J., "Recommending investors for crowdfunding projects," *WWW*, 2014.

[17] Naroditskiy, V., Stein, S., Tonin, M., Tran-Thanh, L., Vlassopoulos, M., and Jennings, N. R., "Referral incentives in crowdfunding," *Second AAAI Conference on Human Computation and Crowdsourcing*, 2014.

[18] Solomon, J., Ma, W., and Wash, R., "Highly Successful Projects Inhibit Coordination on Crowdfunding Sites," .

[19] Mitra, T. and Gilbert, E., "The language that gets people to give: Phrases that predict success on kickstarter," *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, 2014, pp. 49–61.

[20] Xu, A., Yang, X., Rao, H., Fu, W.-T., Huang, S.-W., and Bailey, B. P., "Show me the money!: An analysis of project updates during crowdfunding campaigns," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2014, pp. 591–600.

[21] Hemer, J., "A snapshot on crowdfunding," Tech. rep., Working papers firms and region, 2011.

[22] Haas, P., Blohm, I., and Leimeister, J. M., "An empirical taxonomy of crowdfunding intermediaries," 2014.

[23] Pal, A., Chang, S., and Konstan, J. A., "Evolution of Experts in Question Answering Communities." *ICWSM*, 2012.

[24] Zivkovic, Z., "Improved adaptive Gaussian mixture model for background subtraction," *17th International Conference on Pattern Recognition*, 2004.

[25] Permuter, H., Francos, J., and Jermyn, I., "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognition*, Vol. 39, No. 4, 2006, pp. 695–706.

[26] Manning, C. D., Raghavan, P., and Schütze, H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[27] Kickstarter, "Terms of use," https://www.kickstarter.com/terms-of-use/oct2012, October 2012.