

# Modeling Dynamic Competition on Crowdfunding Markets

Yusan Lin  
The Pennsylvania State University  
University Park, PA 16802, USA  
yusan@psu.edu

Peifeng Yin  
IBM Almaden Research  
San Jose, CA 95120, USA  
peifengy@us.ibm.com

Wang-Chien Lee  
The Pennsylvania State University  
University Park, PA 16802, USA  
wlee@cse.psu.edu

## ABSTRACT

The often fierce competition on crowdfunding markets can significantly affect project success. While various factors have been considered in predicting the success of crowdfunding projects, to the best knowledge of the authors, the phenomenon of competition has not been investigated. In this paper, we study the competition on crowdfunding markets through data analysis, and propose a probabilistic generative model, *Dynamic Market Competition (DMC)* model, to capture the competitiveness of projects in crowdfunding. Through an empirical evaluation using the pledging history of past crowdfunding projects, our approach has shown to capture the competitiveness of projects very well, and significantly outperforms several baseline approaches in predicting the daily collected funds of crowdfunding projects, reducing errors by 31.73% to 45.14%. In addition, our analyses on the correlations between project competitiveness, project design factors, and project success indicate that highly competitive projects, while being winners under various setting of project design factors, are particularly impressive with high pledging goals and high price rewards, comparing to medium and low competitive projects. Finally, the competitiveness of projects learned by DMC is shown to be very useful in applications of predicting final success and days taken to hit pledging goal, reaching 85% accuracy and error of less than 7 days, respectively, with limited information at early pledging stage.

## CCS CONCEPTS

• Information systems → Crowdsourcing; • Social and professional topics → Antitrust and competition; • Applied computing → Business intelligence; Marketing;

## KEYWORDS

market competition; crowdfunding; business competitiveness; probabilistic generative model

## ACM Reference Format:

Yusan Lin, Peifeng Yin, and Wang-Chien Lee. 2018. Modeling Dynamic Competition on Crowdfunding Markets. In *WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3178876.3186170>

## 1 INTRODUCTION

In recent years, crowdfunding has become a new way for entrepreneurs to solicit funding in order to bring their creative ideas and business plans to lives. Crowdfunding platforms, such as Kickstarter, GoFundMe, and Indiegogo, have hence been introduced to help hosting and mediating fundraising campaigns. Numerous reports of funding successes have made such platforms very popular, attracting growing attentions and participations from

entrepreneurs and investors of all kinds.<sup>1,2</sup> While establishing a project is relatively easy on these platforms, it is rather challenging to reach the funding goal in the end. According to statistics released recently by Kickstarter, the largest crowdfunding platform at the time of writing, the average success rate is about 35.83% [26].<sup>3</sup> Many factors may influence the result, e.g., creators' previous success rate/experience [6], promotion on social media [10], language used to describe projects [13], reward bundle design [9] and so on. However, to our best knowledge, the important factor of *competition among projects* in the crowdfunding markets has not received much research attention, even though it has been a topic of research in traditional marketing field [12, 15, 20].

Competition has an important role in affecting a project's success. In a crowdfunding platform, all projects are displayed to backers. Limited resources (available funds in the market) make it a zero-sum game. If one project obtains more backings, other projects may end up with less. In other words, competition exists, either explicit or implicit, among projects. In this work, we aim to build a model to quantify projects' competitiveness in a crowdfunding platform, which can help creators better understand the global picture of market competition and ultimately enhance their chance of successful fund-raising.

Existing works in the literature of marketing mostly quantify the competition in terms of the *competitive pressures*. For example, the degree of product substitutability and the number of competitors are used as measures of the competitive pressures [24]. These existing approaches tend to consider multiple factors that may come into play in competition and employ some heuristics to estimate those factors [3]. However, such an approach relies on careful design of various metrics to capture those factors and sometimes leads to tedious creation of features. At the mean time, it still risks missing some latent features that are not observable. Worse still, the features (and thus the approach) may become constrained by specific domain knowledge, and thus not applicable elsewhere. Therefore, in this work, we aim to propose a model that is general enough to capture different competition scenarios.

To quantify competitiveness, one naive way is to directly use each project's collected funding by assuming the conditions of a *perfect competition*, where "all firms carry the same products" [11, 23]. In reality, projects' contents differ and backers have preferences. As a result, projects with popular content is more likely to receive funding than those only attracting minorities of the population. When quantifying competitiveness, proper incorporation of such a market bias is the first challenge.

To avoid comparing apples to oranges, one intuitive way is to consider only projects of the same content in the same competition, following the scenario of perfect competition. This method is unrealistic. In real crowdfunding platform, it is highly impossible to find two projects with exactly the same content. In general, content similarity is a continuous measure ranging from 0 to 1, rather than

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW 2018, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5639-8/18/04.

<https://doi.org/10.1145/3178876.3186170>

<sup>1</sup><https://www.entrepreneur.com/article/232234>

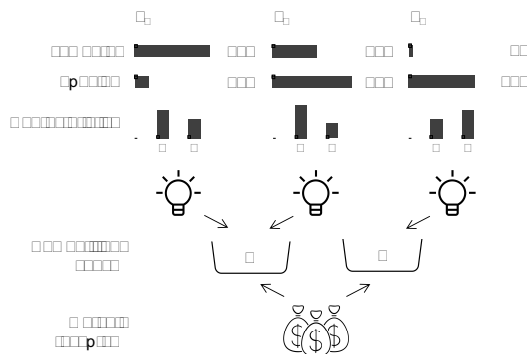
<sup>2</sup><https://www.london.edu/news-and-events/news/start-ups-to-face-competition-as-incumbents-enter-crowdfunding-space>

<sup>3</sup><https://www.kickstarter.com/help/stats>

a binary number 0 or 1. Heuristically, the more similar two projects are, the more likely they will compete for the same resource. The second challenge lies in a proper dealing of content similarity in competition modeling.

Finally, in a crowdfunding platform, the competitiveness of projects changes over time. For example, a new project could attract more backers than a long existing one, due to its *freshness*. On the other hand, a project approaching to its *deadline* may lead to a sense of urgency and a faster increase of funding [14]. Furthermore, a “rich-get-richer” phenomenon suggests that projects close to or exceeding their pledging goals are more likely to attract more new funds [19]. All these dynamic, time-dependent phenomena in crowdfunding competitions need to be considered in the model design.

In this work, we propose a probabilistic model, namely *Dynamic Market Competition (DMC)*, to capture the daily competition in a crowdfunding platform. It assumes there exists  $L$  competition spaces in the market. Each competition space serves as a “virtual arena” for projects to compete against each other. Similar to a probabilistic topic model where latent topics have different word distributions, competition spaces in DMC exhibit different *characteristics* over contents. In everyday competition, each space receives a portion of market resources (funding). Naturally, competition spaces characterized by popular contents would receive more resources than those favoring the non-mainstream ones. Each project also has a vector of *project characteristics* which we model as a probability distribution over the competition spaces, representing the probability of choosing the corresponding space for competition. The project characteristics is the seed of generating the project’s content. Therefore, projects with similar project characteristics have similar contents and thus are more likely to directly compete against each other in the same competition space. The project’s competitiveness is modeled as a real-value number that is dynamically updated. Specifically, its change depends on i) temporal closeness to the deadline and ii) funding progress towards the goal.



**Figure 1: Toy scenario of crowdfunding competition.**

We use a toy example to illustrate our idea of the competition process modeled by DMC. As shown in Figure 1, consider a snapshot of a crowdfunding marketplace (platform), where three active projects  $p_1$ ,  $p_2$  and  $p_3$  are currently pledging for funds. Let  $A$  and  $B$  be two competition spaces in the marketplace and each of them has some amount of funds every day. Every project is randomly drawn to one space based on their project characteristics. Accordingly, projects in the same space “fight” for the allocated funds. In this example,  $p_1$  and  $p_2$  enter the same competition space  $A$  in a certain day and compete against each other. The result of fighting depends on each project’s competitiveness, which is related to temporal

and funding progress. For instance, when considering the temporal progress,  $p_1$  may be more competitive than  $p_2$ , as it may signal a sense of urgency to backers due to the approaching pledging deadline. However, when considering the funding progress,  $p_2$  may be more competitive than  $p_1$ , since it is closer to the pledging goal, i.e., potential backers might view this as a higher chance of success.

Not only can DMC model capture the dynamic competition on market and project competitiveness, it also brings great potentials in terms of crowdfunding analytics, which may provide important insights in the competitiveness of a project and its competitors, and thus enhance the probability of its success. To validate our ideas behind the model, we evaluate the proposed DMC model and demonstrate its potentials by experimenting on two prediction tasks before launching and during pledging: i) predict the final project success, and ii) predict the number of days taken to hit pledging goals. In addition, we exploit the project competitiveness learned by DMC to analyze the correlations between project competitiveness, project design factors, and project success.

The contributions of this paper are summarized as follows.

- (1) To the best knowledge of the authors, this is the first attempt to study the competition among projects on crowdfunding. We conduct analyses that show the presence of competition in crowdfunding markets to reveal various correlated factors, including content, temporal progress, and funding progress.
- (2) We develop a probabilistic generative model, namely *Dynamic Market Competition (DMC)* model, that captures project competitiveness through static and dynamic factors. Compared to several feature-based baseline methods, DMC better predicts projects’ daily collected funds, by reducing errors over these baselines for about 32-45%.
- (3) The DMC model, capturing the dynamic competitiveness of projects very well, can serve as analytics tool to study the correlations between project competitiveness, project design factors, and project success. Our analyses reveal that highly competitive projects, while being winners under various setting of project design factors, are particularly impressive with high pledging goals and high price awards, comparing to medium and low competitive projects.
- (4) Project competitiveness learned from DMC is useful for helping project creators in managing pledging campaigns. We show that project competitiveness can be used along with other baseline features to predict the final project success and the number of days to hit the pledging goal, outperforming baselines by increasing the accuracy for about 7-57% and reducing errors for about 23-41%, respectively.

The rest of this paper is organized as follows. We first review the related work on crowdfunding and the background of competitions in markets in Section 2. We then analyze the data and show the competition in crowdfunding markets in Section 3. We develop Dynamic Market Competition (DMC) model in Section 4. DMC’s ability to predict projects’ daily funds collected compared to other baselines is carried out and reported in Section 5. Insights of correlation between competitiveness, project design, and success rates are discussed, and applications that can be developed based on DMC to serve as project design guidance for creators are evaluated and presented in Section 6. Finally, we conclude this work and lay out future works in Section 7.

## 2 RELATED WORK

In this section, we review the literature relevant to the main aspects of this work, including recent research on crowdfunding and competitions in the crowdfunding markets.

## 2.1 Research on Crowdfunding

Crowdfunding is a unique new type of fundraising platforms, where marketing and competition naturally arise. In these platforms, suppliers, customers, and brokers all influence each others' decisions and behaviors, a scenario known as *peer economy* [29].

Researchers have been studying crowdfunding projects from various angles. One of the most widely explored topics is: *how do different factors affect the success of crowdfunding projects?* Greenberg et al. are one of the pioneers attempting to predict project success on Kickstarter, using only statistics of projects [6]. Xu et al. study how the project updates influence project success [27]. Chung et al. incorporate temporal information and Twitter data to predict project success [4]. Solomon et al. investigate how the timing of donations affects the pledging results on Kickstarter [22]. Lu et al. examine how promotions on social media affect project success [10]. Desai et al. [5] and Mitra et al. [13] focus on the language of project description, and analyze its influence on fund-raising results.

Some works formulate the problem as variant recommendation tasks. An et al. develop a dual recommendation system between backers and projects, considering founders' skill sets, pledging behavior, and projects' geography, growing speed and categories [1]. Rakesh et al. also design a Kickstarter personalized and group recommender, by exploring features in temporal traits, personal traits, geo-location traits, and social network traits [17, 18].

However, these studies view projects as independent entities. There has not been any consideration of the potential *competition* among the projects, which we believe to play an important part in both predicting project success and recommending projects.

## 2.2 Market Competition

Competition is defined as the "rivalry among sellers trying to achieve such goals as increasing profits, market share, and sales volume by varying the elements of the marketing mix: price, product, distribution, and promotion" [25]. The concept of competition has been widely studied in marketing and economics, both theoretically and empirically. However, it has been argued in the marketing literature that the identification of *competitors* is challenging [7, 16]. As a result, competitors in these studies are often arbitrarily given other than automatically identified [12, 15, 20, 21].

A common metric used to capture competition is *competitive pressure*, defined as the pressure a firm receives in developing its own product, while being exposed to creations and actions from other firms in the market [3]. To the best knowledge of the authors, Kim et al. are the first and the only one starting to look into competitions in crowdfunding [8]. However, instead of using competition factors to predict project success, they mainly emphasize on the analysis of the countervailing effects of blockbuster: cannibalization effects and spill-over effects.

Zhang et al. are, to our best knowledge, the first to develop a probabilistic generative model that claims to capture market competition [28]. However, instead of modeling the core of competition, the proposed model learns how brands adopt text and images on social media to establish certain topics. The actual competitiveness of brands has not yet been considered.

In this paper, we propose a new probabilistic generative model that directly models the competition process among projects to learn project competitiveness. It is demonstrated to better predict projects' daily funds collected, final success and number of days taken to hit the goal, so that project creators can better prepare and campaign their projects to bring creative ideas to live.

## 3 COMPETITION ANALYSIS

In this section, we analyze potential factors that may affect a project's competitiveness in the market. We first inspect how the preference and bias the market has towards projects based on their content. We then examine how the static factor, project contents, correlate with their competitiveness. We finally investigate how dynamic factors, including temporal and funding progress, influence projects' competitiveness. Results from these analyses shed light on our model design.

To carry out the analyses, we collect a rich dataset that captures the vibrant changes in crowdfunding market from Kickstarter, currently one of the most prominent platforms. Since founded in April 2009, Kickstarter has helped 103K project creators pledge over \$2.29 billions of dollars. 10 million backers have been involved in funding those projects, whereas 29 million backings have been made.<sup>4</sup> Being such an enormous platform for brewing creative ideas, we choose it as the data source of our study. We acquire data of a three-month period, from December 15<sup>th</sup>, 2013 to March 14<sup>th</sup>, 2014. The statistics of the dataset are summarized in Table 1.

**Daily distribution of market funds:** Since the market and backers have bias towards certain types of projects with more popular contents, we are interested in how funds are distributed from the market. We define the "type" of projects by clustering. Particularly we use k-means to cluster projects into five clusters based on their contents. We then observe the percentage of funds received from the market by each cluster of projects every day.

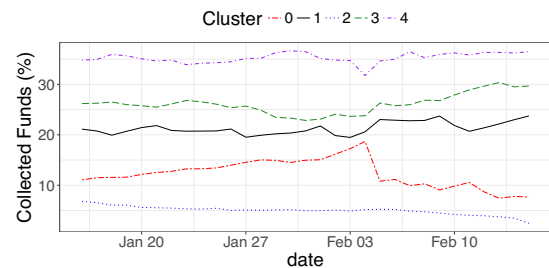


Figure 2: Daily market share of each project cluster

As Figure 2 shows, the clusters' order in received proportion of funds stay unchanged throughout the observed duration, showing a strong bias market has towards project contents. Also, the proportion of funds each project type (i.e., cluster) receives is rather stable throughout the days. This indicates that the market's distribution of funds to types of projects are relatively static.

**Project content v.s. final funding collected:** Knowing the market has bias towards project contents, now we analyze how the project contents correlate with their competitiveness, and whether project contents solely dictate project competitiveness. We approach this by checking whether the collected funds within the same project type are less variant than that in the overall market.

Following the previous analysis, we adopt the same clusters to identify types of project contents. Figure 3 shows the standard deviations (STD) of collected funds normalized by the total collected funds for each cluster of projects. As shown, the within-cluster standard deviations all exceed the overall standard deviation, indicating the unbalanced allocation of funding among similar projects. This scenario thus confirms the existence of competition.

**Dynamic changes v.s. project competitiveness:** At last, we analyze the dynamics of projects' competitiveness. For each project,

<sup>4</sup><https://www.kickstarter.com/help/stats>



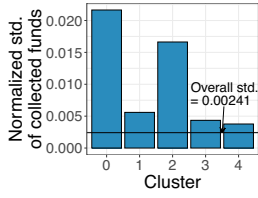


Figure 3: Project clusters v.s. STD of funds collected

Table 1: Statistics of the Kick-starter dataset

	Projects	Rewards	Raised
Success	1,399	15,472	\$27,534K
Fail	1,705	14,866	\$3,758K
Total	3,104	30,338	\$31,293K

we calculate its funding and temporal progress during its active period. Specifically, the funding progress is defined as the ratio of collected funding amount to the pledging goal, whereas the temporal progress is defined as the ratio of active days to the total number of funding days (i.e., from the starting date to the deadline).

Formally, let  $pledged(i, t)$  denote the funding amount a project  $i$  has collected until time  $t$ , and  $goal(i)$  be the funding goal project  $i$  sets. Given project  $i$  at time  $t$ , its funding progress is defined as follows.

$$r(i, t) = \frac{pledged(i, t) - goal(i)}{goal(i)} \quad (1)$$

Similarly, the formal definition of temporal progress is given below.

$$d(i, t) = \frac{t - start\_date(i)}{duration(i)} \quad (2)$$

For funding/temporal progress in each day of the active period, there is a corresponding *daily gain*, i.e., funding received daily. We plot the curve of average daily gain in Figures 4(a) and 4(b).

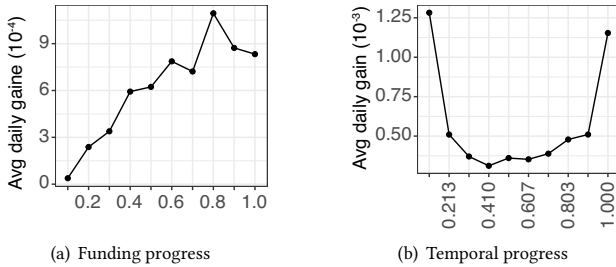


Figure 4: Projects' daily gain v.s. two types of progresses

As shown, the daily gain has a positive linear relationship with the funding progress, indicating that while the pledging progress approaches closer to the pledging goal, a project receives more daily funds. On the other hand, the daily gain has a bimodal relationship with the temporal progress, suggesting that projects receive more funds at the beginning and the end of the pledging period.

#### 4 DYNAMIC MARKET COMPETITION MODEL

A project's fund pledging performance can reflect its competitiveness against others on the market. In this section, we develop a model that captures a project's competitiveness over time based on its pledging result compared to others and the status of the market.

As mentioned earlier in Section 1, there does not exist *perfect competition* on the crowdfunding market. Instead, we assume there implicitly exist multiple *competition spaces* where similar projects compete with each other for funding. Accordingly, we conceptually describe the competition in a crowdfunding platform as follows. Every day, thousands of projects on the crowdfunding platform

Table 2: Symbol definitions

Symbol	Definition
$W$	number of project features
$N$	number of total projects
$T$	number of total days
$M$	amount of funds on market
$L$	number of competition spaces
$a$	funding progress impact
$b, c$	temporal progress impact
$\alpha, \delta$	Dirichlet priors
$v_{i,j}$	latent topic of project $i$ on project feature $j$
$w_{i,j}$	project $i$ 's feature $j$
$\phi_l$	topic distribution over competition space $l$
$\lambda_l$	market share on latent competition space $l$
$\eta_{i,l}$	preference of project $i$ on latent competition space $l$
$x_{i,t}$	competitiveness of project $i$ on day $t$
$s_{i,t}^p$	assigned competition space of project $i$ on day $t$
$s_{t,m}^f$	assigned competition space of fund $m$ on day $t$
$\Delta_{i,t}$	daily gain of project $i$ on day $t$
$\theta_i$	fund distribution on project $i$
$z_{t,m}$	assigned project of fund $m$ on day $t$
$g_{i,t}$	funds received by project $i$ on day $t$

compete against each other to fight for financial supports from backers. The more *competitive* a project is, the more likely it is to attract backers' attentions and further receive funds from them. Imagine that every project has its own probabilistic rule to randomly pick the competition space. Such rule is determined according to characteristics of their products. Therefore similar projects are likely to choose the same competition space to directly compete against each other. After a day of combat, each project takes home with their collected funds from the backers, which reflect their *competitiveness* of the day. The projects then enter the market again the next day with the updated competitiveness, and compete with other projects in a newly drawn competition space over the available resource.<sup>5</sup> This process repeats daily for a project until its pledging deadline.

In the following, we first introduce the model parameters and definitions of concepts by linking them to the competition process in order to ease the discussion of the generative process in DMC. For clarity, we summarize all symbols in Table 2.

First of all, we denote the resource (i.e., amount of funds) available for all projects in the market in a day as  $M$ , which can be seen as a big bucket containing  $M$  coins aggregated from all backers in the day, ready for distributing to projects as their financial supports. The resource is distributed over  $L$  competition spaces based on a variable  $\lambda$ , namely *market share distribution*, which is represented as a distribution vector. As such, the  $l^{\text{th}}$  competition space is allocated with a fixed share of the resources  $M \cdot \lambda_l$  from the market's resource bucket. We formally define the market and the competition space as follows.

**DEFINITION 1.** A *market* is the complete environment of observation, i.e., the observed crowdfunding platform under the context of this paper. A market consists of  $L$  **competition spaces**,  $s_1, s_2, \dots, s_L$ . Each *competition space* involves similar projects that compete against each other to fight for resources. The market distributes resources  $M$  to the  $L$  competition spaces based on a **market share distribution**  $\lambda$ , where the  $l^{\text{th}}$  competition space receives resources of  $\lambda_l M$ , and  $\sum_l \lambda_l M = M$ .

<sup>5</sup>Generally any time unit may be used in the recurring process. In this problem, we choose to use one day to match the actual backing cycle.

For a project  $p_i$ , it randomly enters a competition space in accordance with its *characteristic vector*,  $\eta_i$ , which is represented as an  $L$ -dimensional distribution determined by the product characteristics of  $p_i$ . Probabilistically, the project fights for funds allocated to all its corresponding competition spaces. For a given competition space  $s_l$ ,  $p_i$  exploits the power of its competitiveness  $x_i$  to fight for the resource allocated to  $s_l$ .

Taking a complementary angle of *fund allocation* to see the competition process, it's like there is a *pseudo backer* making funding decisions on behalf of all backers. For a given fund  $m \in M$  on the market, the pseudo backer first decides which of the  $L$  spaces to contribute to, based on the market share distribution  $\lambda$  probabilistically. It then decides a project to back based on the competitiveness of projects in the corresponding competition space, where projects with higher competitiveness are more likely to receive funds. Based on the above, we formally introduce the generative process of the DMC model, followed by the design of parameter inference.

#### 4.1 Generative Process

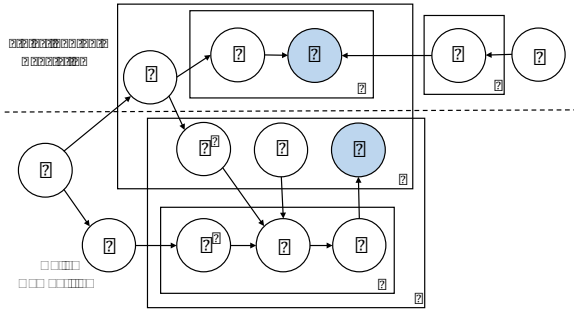


Figure 5: Dynamic Market Competition (DMC) model

Figure 5 shows the graphical representation of the proposed model, namely, *Dynamic Market Competition (DMC)* model. As can be seen, the generative process of DMC consists of two stages: (1) project content generation, and (2) daily project competition and market fund distribution.

**4.1.1 Project content generation.** We assume that a project is characterized by a data generation variable,  $\eta$ , a probabilistic distribution over the  $L$  topics in project content, which is an aggregation of categories, descriptions, rewards, price, duration, and pledging goals. The graphical model for this stage is depicted in the upper part of Figure 5. For simplicity, we model the project content as a bag-of-words feature vector of size  $W$  and follow the well-received *latent Dirichlet allocation (LDA)* [2]. We do not elaborate the LDA model but note that  $v$  is the topic variable dictating project content  $w$ . Moreover, we term the variable  $\eta$  as *project characteristics*, which plays an important role in Stage 2 to determine the competition space in a daily basis.

Note that as the content of a project usually stays unchanged throughout the pledging duration after its launch, we do not update  $\eta$  through out the second stage, i.e., the project competition process.

**4.1.2 Project competition process.** In the second stage, we model the daily competition among projects in terms of fund distribution (by pseudo backer). The bottom part of Figure 5 shows the graphical representation. Specifically, this stage is further divided into two steps: *project competition space setup* and *market fund allocation*. The former determines which competition space a project enters to compete based on its project characteristics generated in Stage 1, and the latter describes how a fund is distributed to a project.

**Project competition space setup:** The competition spaces are corresponding to the  $L$  topics as discussed in Stage 1. To prepare for the daily competition in each competition space on each day  $t$ , this step randomly assigns every active project  $p_i (i = 1..N)$  to one of the competition spaces,  $s_{i,t}^p$ , based on its project characteristics  $\eta_i$ . Formally, for each of the  $T$  days, for each of the  $N$  projects, a participating competition space of the  $i^{\text{th}}$  project on the  $t^{\text{th}}$  day, i.e.,  $s_{i,t}^p$ , is drawn as follows.

$$s_{i,t}^p \sim \text{Multinomial}(\eta_i)$$

Meanwhile, for each project  $p_i$ , an initial competitiveness  $x_{i,0}$  that dictates its future competitiveness is drawn from the Gaussian distribution (see below) before any funding progress and temporal progress is made.

$$x_{i,0} \sim \mathcal{N}(\mu_0, \sigma)$$

Given the initial competitiveness  $x_{i,0}$  of project  $p_i$ , for each day  $t$  during its active pledging period, the daily competitiveness  $x_{i,t}$  is affected by both its funding progress and temporal progress, as defined in Eq. (1) and Eq. (2), respectively, in Section 3.

Recall the analytical result in Figure 4(a). The projects' daily gain is a linear function of the funding progress. Thus we use a linear function  $a \cdot r_{i,t}$  to express it, where  $r_{i,t}$  is the funding progress of the  $i^{\text{th}}$  project at day  $t$ . As shown in Figure 4(b), the projects' daily gain is a quadratic function of the time progress. Hence we use  $b \cdot d_{i,t}^2 + c \cdot d_{i,t}$  to express it, where  $d_{i,t}$  is the corresponding temporal progress. Putting them together, a project  $p_i$ 's competitiveness at day  $t$  is a mixture of its initial competitiveness  $x_{i,0}$ , the funding progress and temporal progress.

$$x_{i,t} = \underbrace{x_{i,0}}_{\text{initial competitiveness}} + \underbrace{a \cdot r_{i,t}}_{\text{funding progress}} + \underbrace{b \cdot d_{i,t}^2 + c \cdot d_{i,t}}_{\text{temporal progress}}$$

**Market fund distribution:** After the setup of competition space and each project's competitiveness are all settled, in this step, the pseudo backer first distributes the market funds to a competition space, and then further distribute it to a project in the competition space iteratively. The generative process is described as follows.

We first obtain the market share distribution  $\lambda$ , which is a  $L$ -dimensional probability distribution over the competition spaces, from the Dirichlet distribution

$$\lambda \sim \text{Dirichlet}(\alpha)$$

Based on our analysis in Section 3, we assume that the market share is stable in allocating funds to different competition spaces. Hence  $\lambda$  stays the same throughout all days  $T$ .

Accordingly, for each day  $t$ , a fund  $m \in M$  is drawn into a competition space  $s_{t,m}^f$  as follows.

$$s_{t,m}^f \sim \text{Multinomial}(\lambda)$$

Next, a probability distribution  $\theta$  over all of the projects in the space  $s_{t,m}^f$  is generated as follows.

$$\theta \sim \text{Dirichlet}(\{x_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = s_{t,m}^f}\}) \quad (3)$$

where  $\mathbf{1}_{s_{i,t}^p = s_{t,m}^f}$  constrained that only projects assigned to the same competition space as the fund  $m$  are considered for the fund distribution. Also, the competitiveness  $x_{i,t}$  in Eq. (3) forces fund  $m$  to land on highly competitive projects with higher probability.

Therefore, a project  $z_{t,m}$  is chosen based on the distribution  $\theta$  to receive the fund  $m$  on day  $t$  as below.

$$z_{t,m} \sim \text{Multinomial}(\theta)$$

The overall generative process of the project competition in Stage 2 including the project competition space assignment and market fund distribution is summarized in Algorithm 1.

---

**Algorithm 1** Generative process for competition model.

---

```

1:  $\lambda \sim \text{Dirichlet}(\alpha)$ 
2:  $a, b, c \sim \mathcal{N}(0, \sigma)$ 
3: for each project  $i$  do
4:    $x_{i,0} \leftarrow \mathcal{N}(\mu_0, \sigma)$ 
5: end for
6: for each day  $t$  do
7:   for each project  $i$  do
8:     update  $r_{i,t}, d_{i,t}$  according to  $g_{i,t-1}$  and  $t$ 
9:      $x_{i,t} \leftarrow x_{i,0} + a \cdot r_{i,t} + b \cdot d_{i,t}^2 + c \cdot d_{i,t}$ 
10:     $s_{i,t}^p \sim \text{Multinomial}(\eta_i)$ 
11:   end for
12:   for each available fund  $m$  do
13:      $s_{t,m}^f \sim \text{Multinomial}(\lambda)$ 
14:      $\theta \sim \text{Dirichlet}(\{x_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = s_{t,m}^f}\})$ 
15:      $z_{t,m} \sim \text{Multinomial}(\theta)$ 
16:   end for
17:    $g_{i,t} \leftarrow \sum_m \mathbf{1}_{z_{t,m}=i}$ 
18: end for

```

---

## 4.2 Maximum Likelihood Estimation

We consider Maximum Likelihood Estimation (MLE) for obtaining the model parameters  $\Theta = \{X, S^p, Z, S^f, \lambda, a, b, c\}$ . Note that since project content generation, the first stage of the model, is basically the well known LDA, we only present the likelihood estimation and parameter inference of the second stage, project competition process, for the sake of space saving.

According to the DMC model, we express the log-likelihood for the parameters in  $\Theta$  based on the observations as follows.

$$\begin{aligned} & \log P(X, S^p, Z, S^f, \lambda, a, b, c | G, \alpha, \eta, \mu_0, \sigma) \\ & \propto \log P(S^p, Z, S^f | X, G, \lambda, \eta) + \log P(X | \mu_0, \sigma) + \log P(\lambda | \alpha) \quad (4) \\ & \quad + \log P(a | \sigma) + \log P(b | \sigma) + \log P(c | \sigma) \end{aligned}$$

Note that besides  $X$ , we also introduce Gaussian prior distribution over dynamic parameters  $a, b, c$ . They share the same variance  $\sigma$  with  $X$ , but their mean is 0, i.e., the prior of the competitiveness is static.

The first term can be further written as:

$$\begin{aligned} & \log P(S^p, Z, S^f | X, G, \lambda, \eta) = \sum_t \log P(S_{t,\cdot}^f, S_{t,\cdot}^p, Z_{t,\cdot} | X_{t,\cdot}, G_{t,\cdot}, \lambda, \eta) \\ & = \sum_t \log P(S_{t,\cdot}^f | \lambda, G_{t,\cdot}, S_{t,\cdot}^p) + \log P(Z_{t,\cdot} | X_{t,\cdot}, S_{t,\cdot}^p, G_{t,\cdot}) + \log P(S_{t,\cdot}^p | \eta) \end{aligned}$$

Particularly, given  $S_{t,\cdot}^p$  and  $G_{t,\cdot}$ , the probability for funding allocation can be expressed as:

$$P(S_{t,\cdot}^f | \lambda, G_{t,\cdot}, S_{t,\cdot}^p) = \prod_l \lambda_l^{\sum_i g_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = l}}$$

Therefore, the whole competition process is as follows.

$$\begin{aligned} P(Z_{t,\cdot} | X_{t,\cdot}, S_{t,\cdot}^p, G_{t,\cdot}) &= \prod_m P(Z_{t,m} | X_{t,\cdot}, S_{t,\cdot}^p, G_{t,\cdot}) \\ &= \prod_l \frac{\prod_i x_{i,t}^{g_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = l}}}{(\sum_j x_{j,t} \cdot \mathbf{1}_{s_{j,t}^p = l})^{\sum_k g_{k,t} \cdot \mathbf{1}_{s_{k,t}^p = l}}} \end{aligned}$$

To sum up all of the above equations, the objective function can therefore be defined as:

$$\begin{aligned} \mathcal{L}(\Theta) &= \log P(S^f | \lambda, G, Y) + \log P(Z | X, S^p, G) + \log P(X | \mu_0, \sigma) + \log P(\lambda | \alpha) \\ &= \sum_{t,l} \left( \sum_i g_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = l} \right) \log \lambda_l + \sum_{t,i} g_{i,t} \log x_{i,t} \\ & \quad - \sum_{t,l} \left( \sum_i g_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = l} \right) \log \sum_j x_{j,t} \cdot \mathbf{1}_{s_{j,t}^p = l} \quad (5) \\ & \quad - \frac{\sum_i (x_{i,0} - \mu_0)^2}{2\sigma^2} + \sum_l (\alpha - 1) \log \lambda_l - \frac{a^2}{2\sigma^2} - \frac{b^2}{2\sigma^2} - \frac{c^2}{2\sigma^2} \end{aligned}$$

## 4.3 Parameter Inference

To learn the parameters  $\Theta$  that maximize the log-likelihood defined as Eq. (5), we leverage Expectation Maximization (EM) algorithm. We design the E- and M-steps as follow.

**E-step:** As seen, the hidden variable  $S^p$  is the key for the log-likelihood shown in Eq. (5). Therefore, we compute the expected value, or exactly, the  $E(\mathbf{1}_{y_{i,t}=l})$ :

$$E(\mathbf{1}_{s_{i,t}^p = l}) = 1 \cdot P(s_{i,t}^p = l | \eta_i) + 0 \cdot P(s_{i,t}^p \neq l | \eta_i) = \eta_{i,l} \quad (6)$$

The expected value of  $S^p$  is therefore fixed and stays unchanged throughout iterations.

**M-step:** With expected value of  $\mathbf{1}_{s_{i,t}^p = l}$ , we obtain parameter values by maximizing the log-likelihood conditioned on such expected value.

$$\Theta^* = \arg \max_{\Theta} \mathcal{L}_{E(\mathbf{1}_{s_{i,t}^p = l})}(\Theta) \quad (7)$$

By setting  $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$  and constraining that  $\sum_l \lambda_l = 1$ , we obtain:

$$\lambda_l = \frac{\alpha - 1 + \sum_{i,l} g_{i,t} \cdot \eta_{i,l}}{L \cdot (\alpha - 1) + \sum_{t,l} g_{i,t} \cdot \eta_{i,l}} \quad (8)$$

For  $X_{i,0}$ , we can compute partial derivation via  $x_{i,t}$  and note that  $\frac{\partial x_{i,t}}{\partial x_{i,0}} = 1$ , we have:

$$\frac{\partial \mathcal{L}}{\partial x_{i,0}} = \frac{\sum_t g_{i,t}}{x_{i,t}} - \sum_{t,l} \frac{\sum_j g_{j,t} \cdot \eta_{j,l}}{\sum_j x_{j,t} \cdot \eta_{j,l}} \cdot \eta_{i,l} - \frac{x_{i,0} - \mu_0}{\sigma^2} \quad (9)$$

Using similar chain rule, we obtain the partial derivation for  $a, b$  and  $c$  as follows.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial a} &= \sum_{t,i} \frac{g_{i,t}}{x_{i,t}} \cdot r_{i,t} - \sum_{t,l} \frac{\sum_j g_{j,t} \cdot \eta_{j,l}}{\sum_j x_{j,t} \cdot \eta_{j,l}} \cdot \sum_i \eta_{i,l} \cdot r_{i,t} - \frac{a}{\sigma^2} \\ \frac{\partial \mathcal{L}}{\partial b} &= \sum_{t,i} \frac{g_{i,t}}{x_{i,t}} \cdot d_{i,t}^2 - \sum_{t,l} \frac{\sum_j g_{j,t} \cdot \eta_{j,l}}{\sum_j x_{j,t} \cdot \eta_{j,l}} \cdot \sum_i \eta_{i,l} \cdot d_{i,t}^2 - \frac{b}{\sigma^2} \quad (10) \\ \frac{\partial \mathcal{L}}{\partial c} &= \sum_{t,i} \frac{g_{i,t}}{x_{i,t}} \cdot d_{i,t} - \sum_{t,l} \frac{\sum_j g_{j,t} \cdot \eta_{j,l}}{\sum_j x_{j,t} \cdot \eta_{j,l}} \cdot \sum_i \eta_{i,l} \cdot d_{i,t} - \frac{c}{\sigma^2} \end{aligned}$$

Note that there is no explicit solution for  $X_{i,0}$  and  $a, b, c$ . Hence we use gradient descent to indirectly find them. More specifically, we first randomly initialize  $X_{i,0}$  and  $a, b, c$ . We then use Eq. (9) and

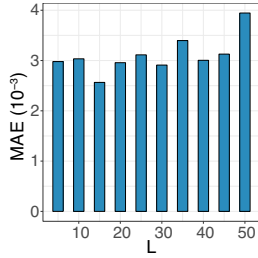


Figure 6: DMC’s performance of daily funds prediction with various L

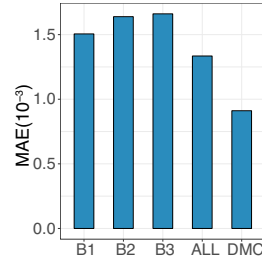


Figure 7: DMC and baselines’ performance of daily funds prediction

Eq. (10) as gradients in each iteration to update the parameters, and the iterations are repeated until convergence.

## 5 MODEL EVALUATION

In this section, we evaluate the quality of the competitiveness learned by DMC. A good set of project competitiveness should be able to accurately reflect the projects’ performance in earning funds. Therefore, we use *daily collected fund prediction* to assess the effectiveness of DMC.

### 5.1 Experiment Setup

We use the three-month Kickstarter data, as introduced in Section 3, for evaluation. Particularly, for each time point (day), we predict the gained funding for each project using all previous data for training. Accordingly, we report the average difference between our prediction and the ground truth, i.e., Mean Absolute Error (MAE), as defined below.

$$MAE = \frac{1}{T} \sum_t \frac{1}{N_t} \sum_i |\hat{g}_{i,t} - g_{i,t}| \quad (11)$$

where  $\hat{g}_{i,t}$  is the predicted funds received by project  $i$  on day  $t$ , and  $g_{i,t}$  is the actual funds received.

### 5.2 Number of Competition Space

In DMC, one of the important hyper-parameters is  $L$ , the number of competition spaces. In the first experiment, we examine how DMC performs with varied  $L$ . We approach this by running simulations. Specifically, let  $z_{t,m}$  be the project a fund  $m$  available in day  $t$  lands on. The project  $z_{t,m}$  can be empirically derived as follows.

$$z_{t,m} = \arg \max_{i^*} \frac{x_{i^*,t}}{\sum_i x_{i,t} \cdot \mathbf{1}_{s_{i,t}^p = s_{m,t}^f}} \quad (12)$$

where  $s_{i,t}^p$  is the space a project  $i$  enters in day  $t$ , and  $s_{m,t}^f$  is the space the fund  $m$  lands on in day  $t$ . The funds received by project  $i$  on day  $t$  is then obtained as follows.

$$\hat{g}_{i,t} = \sum_m \mathbf{1}_{z_{t,m}=i} \quad (13)$$

For each value of  $L$ , we run the simulations for 10 rounds to obtain the average MAE. As shown in Figure 6, DMC achieves the lowest MAE when  $L = 15$ . We thus set the number of competition spaces as 15 for the rest of the evaluation.

### 5.3 Comparison with Baselines

We compare DMC’s effectiveness in predicting daily collected funds with baseline methods that exploit selective features, as summarized in Table 3, to train regression models.<sup>6</sup> The first baseline (B1) consists of *basic project features* proposed for crowdfunding project

<sup>6</sup>For baseline features, we choose those that can be *internally generated from Kickstarter*, without resorting to external resources, such as Twitter and Facebook.

Table 3: Baselines

Baseline	Features
B1: Basic project features	Success ratio of the creator [17] SMOG grade of reward description [4] SMOG grade of main page description [4] Number of sentences in reward description [4] Number of sentences in project description [4] Connected to twitter [6]
B2: Blockbuster effect [8]	Number of blockbusters inside cluster Number of blockbusters outside cluster Number of projects inside cluster Number of projects outside cluster
B3: Competitive pressure	Product similarity pressure Pricing pressure Funding progress pressure Temporal progress pressure

success prediction (introduced in Section 2); the second baseline (B2) consists of *blockbuster effect features* proposed by Kim et al. to capture the competition brought by blockbuster projects [8]; the third baseline (B3) consists of *competitive pressure features* based on heuristics proposed to capture competitive pressures [3]. Those competitive features can be classified in four types: 1) *similarity pressure* measured by the average similarity of projects to a given project  $i$  on a given day; 2) *price pressure* measured by the price of a given project  $i$  to the average price of all other active projects on a given day; 3) *funding progress pressure* measured by the funding progress (see Eq. (1)) of a given project  $i$  to the average funding progress of all other active projects on a given day; and 4) *temporal progress pressure* measured by the temporal progress (see Eq. (2)) of a given project  $i$  to the average temporal progress of all other active projects on a given day. To fully exploit the strengths of these baseline features, we also learn an additional regression model (denoted as ALL) that exploits all B1, B2 and B3 features.

As mentioned, we formulate the prediction task as a regression problem for all baselines. The features are used as independent variables, and the daily funds collected by each project every day are used as the dependent variable for a linear regression model. As shown in Figure 7, DMC significantly outperforms all baselines, B1, B2, B3 and ALL in terms of reduced MAE by 39.47%, 44.41%, 45.14%, 31.73%, respectively.

## 6 COMPETITIVENESS ANALYSIS AND APPLICATIONS

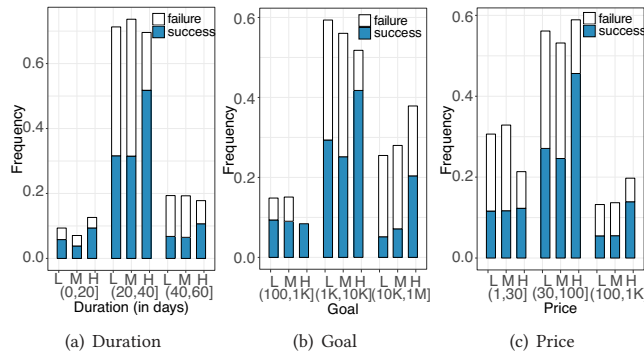
Through the high accuracy in predicting projects’ daily collected funds, we have demonstrated the effectiveness of DMC. In this section, we take a closer look at the correlations of project competitiveness to some project design factors and the project success. In addition, we evaluate two applications by exploring the project competitiveness learned by DMC. These experiments are particularly useful for project creators to better understand their project competitiveness in early stage and assist them to make appropriate adjustments.

### 6.1 Competitiveness and Project Design Factors

Project characteristics (probabilistically) decides which competition space a project enters to compete, that ultimately has an impact on its success. Main factors in project design that affect project characteristics include *pledging duration*, *pledging goal*, and *reward price range*. We thus examine how projects of different competitiveness would make different design choices of these factors, and how that correlates with their final success rate. To proceed, we first classify all projects by their level of competitiveness in terms of percentiles: *Low (L)* indicates competitiveness lower than the 10<sup>th</sup> percentile;



High (H) indicates competitiveness higher than the 90<sup>th</sup> percentile; and Medium (M) for the rest.

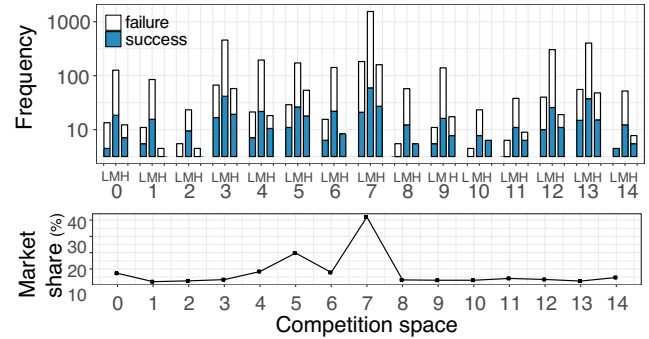


**Figure 8: Project design v.s. project success rate for different project competitiveness**

**Pledging duration:** Creators decide the pledging durations of their projects before launching. Figure 8(a) shows the frequency of low, medium and high competitive projects under different pledging durations. Note that the duration is expressed by percentage of projects under certain duration range out of the corresponding competitiveness level, e.g., 5.54%, 66% and 28.45% of high competitive projects are in (1,20], (20,40] and (40,60] ranges, respectively. As shown, regardless of competitiveness levels, least percentage of projects choose 1-20 days for their pledging period. However, these projects generally have good success rates (Low: 61.67%, Medium: 54.64%, High: 79.34%), while high competitive projects prevail. On the other hand, for pledging durations of 20-40 days, which are designated by most projects in all competitive levels, the success rate of high competitive projects is obviously much higher than the other projects (Low: 52.12%, Medium: 57.22%, High: 74.02%). Finally, for pledging period of 40-60 days, the percentage of projects from each competitiveness level are pretty even but again the high competitive projects are much more successful than the other projects (Low: 32.76%, Medium: 32.80%, High: 59.16%). In summary, project success is highly correlated to the project competitiveness, regardless the pledging period.

**Pledging goal:** Creators set the pledging goals of their projects to disclose their desired amount of funds to collect. Figure 8(b) shows the frequencies of projects with different competitiveness levels under varied pledging goals. As shown, for projects setting pledging goals less than \$1000 USD, their success rates are the highest (Low: 63.04%, Medium: 59.89%, High: 100.00%). For projects setting their goals above \$1000 USD, their success rates are lower (for goals (\$1K, \$10K], Low: 49.46%, Medium: 44.92%, High: 80.62%; for goals (\$10K, \$1M]: Low: 20.25%, Medium: 25.50%, High: 53.84%). However, even when setting the goal higher than \$10K USD, high competitive projects are still able to reach success rates over 50%, while others can barely exceed 25%. This indicates that even with high and hard to reach pledging goal set, high competitive projects are still able to succeed more than others.

**Reward price range:** Creators use rewards which contain various items under different prices to give back to (or attract) backers who contribute to the pledging. We use the medium of all reward prices in a project as its representative price. Figure 8(c) shows the frequencies of projects with different competitiveness levels, under increased price ranges. As shown, in all price ranges, high competitive projects have the highest success rates (Prices (\$1,\$30]:



**Figure 9: Project distribution (upper) and market distribution (bottom) over competition space**

Low: 37.89%, Medium: 35.58%, High: 57.57%; prices (\$30, \$100]: Low: 48.27%, Medium: 46.28%, High: 77.47%; prices (\$100, \$1K]: Low: 41.46%, Medium: 40.41%, High: 70.49%). We observe that, although low and medium competitive projects have overall lower success rates, using price between \$30 and \$100 USD is a relatively safe strategy. This may be due to the fact that with non-high competitiveness, backers are not willing to pay for high-priced rewards, and low-priced rewards are hard to accumulate enough funds to reach the pledging goal. On the contrary, it's interesting to find that high competitive projects with higher-priced rewards are more successful than those with rewards in the lowest price range.

**Competition space:** In addition to the design factors discussed above, project contents are also involved in determining the participating competition space of a project. Some competition spaces are more popular due to the market share distribution, and some prevalent competition spaces have more projects participating because the characteristics of these project are common. In this analysis, we show the distribution of projects in different competitiveness levels for each competition space. As Figure 9 shows, competition space #7 has the most projects participating (40.70%). It also has the most funds received (41.93%). However, it does not have the highest success rate (37.35% compared to 45.06% overall). On the other hand, while space #14 has very few projects involved (1.2%), i.e., a niche market, it has the highest success rates among all competition spaces (50%). Taking a closer look into projects in #14, we observe that projects in this space mainly focus on performing arts such as fashion, dance, and musical shows. Although these projects receive relatively low market share (4.74%), they are more likely to succeed due to the less competition from each other and potentially the loyal backers.

Intuitively, in every competition space, projects with high competitiveness should have higher success rates, and vice versa, just like what spaces #8 and #10 exhibit. However, the high competitive projects in spaces #1 and #2 have success rate of 0%. Case studies reveal some interesting findings. An example in space #2 (where there is limited market share, 2.69%, and most project contents are movie- and film-related) is the project, *Blind Faith Feature Film Project*, which has high competitiveness but failed to reach the pledging goal. When looking into the project design, we find that it sets an extremely high goal of \$4M. On the other hand, a project with medium competitiveness in the same space, *MPC Pre-release Limited Edition Playing Cards*, which is a tabletop game project, sets a lower goal of \$1,500 and results in pledging success. Such result may be due to two reasons: 1) even with a high competitiveness, when a project has a pledging goal way higher than the others in the same competition space, it is hard to achieve its goal; 2) if a



project is unique in certain aspects (such as its product) among all the projects in the same competition space, it may be more likely to achieve the goal.

## 6.2 Applications

In this section, we demonstrate that via the daily project competitiveness identified, DMC may help in two useful applications: 1) predict the final success of a project, and 2) predict the number of days a project may take to hit the pledging goal. In practice, these predictions may help the creators to better understand their chances of reaching the pledging goal before the campaign ends, hence providing guidance for the creators to adjust their design of projects. In the following, we incorporate the project competitiveness (learned by DMC and thus denoted as DMC), along with the baseline features introduced earlier in Table 3, to train classifiers and regression models for the aforementioned applications.

**Application 1. Predict the final pledging result:** Knowing beforehand the estimated project success rate can inform creators whether the project is on track or in danger to succeed. We formulate this task as a classification problem, with binary classes of *success* or *failure*. Using random forest as the classifier, the prediction is done on a daily basis with accumulative information, i.e., a project's success prediction done on day  $t$  uses all information from day 1 to  $t - 1$ . The classifiers are evaluated with 10-fold cross validation. Figure 10 shows the performance of the baseline features and DMC in both daily basis and average. As shown, when using the project competitiveness (DMC) along with all other features in the baselines, the predictiveness reaches the highest in all performance metrics: accuracy, precision and recall. The average accuracy of DMC achieves over 85%, outperforming B1, B2, B3, and ALL by 22.66%, 56.72%, 9.96%, 7.37%, respectively. Moreover, the prediction performance throughout all days in the pledging period are very steady, achieving 83% in the worst day. This shows that even with limited information in the early stage of project pledging, one can predict the project success as well as when more information becomes available while the pledging goes on. With the help of DMC, one can better foresee the projects' final success.

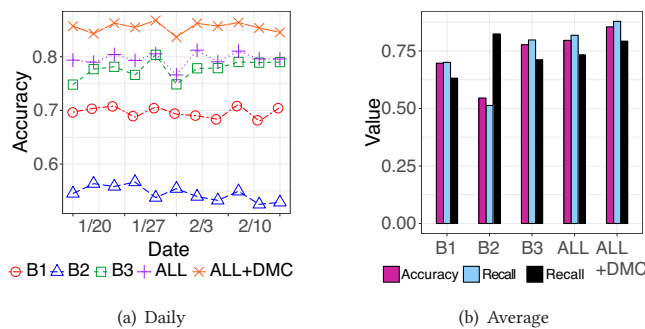


Figure 10: Performance of project success prediction

**Application 2. Predict the number of days to hit pledging goal:** Besides the final pledging goal, achieving the goals fast is also desirable for project creators. We hence demonstrate the power of project competitiveness learned by DMC in predicting the number of days taken for a project to hit its pledging goal. Because the target is a numerical variable, we formulate this task as a *regression* problem, where one project's input on day  $t$  include all information available from day 1 to  $t - 1$ , while the output is the number of days to hit the goal. The models are evaluated with 10-fold cross

validation. Figure 11 shows the performance of baselines and DMC in both daily basis and average. As shown, when the competitiveness learned by DMC are combined with all the baseline features, the MAE of predicting days taken to hit project goals is the lowest on all days and on average. The MAE of DMC in this experiment is about 7 days, which outperforms B1, B2, B3, and ALL by 26.09%, 40.54%, 40.27%, 22.51%, respectively. This shows that with the help of DMC, one can better predict the amount of time to reach the pledging goal.

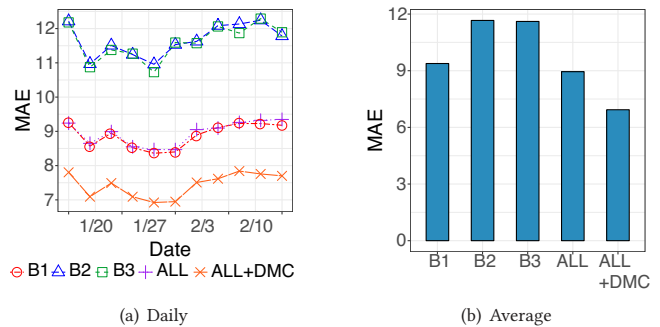


Figure 11: Performance of days to hit goal

The above experimental results validate DMC's effectiveness in predicting the final project success, and the number of days to hit the pledging goal. With DMC's help, creators are able to foresee the potential pledging result based on their current project characteristics and competitiveness, and can adjust accordingly to enhance success rate.

## 7 CONCLUSION AND FUTURE WORKS

Competition is a prevalent phenomenon in various types of marketing, and crowdfunding is no exception. Due to a lack of relevant research, in this work, we analyze the presence of competition in crowdfunding, propose a probabilistic generative model, Dynamic Market Competition (DMC) model, that captures the project competitiveness, and further predicts project success.

We analyze the competition on the crowdfunding market and unveil affecting factors, such as project contents, temporal progress and funding progress. We then develop the DMC model that models the crowdfunding market as a competition scenario. With the learned model, we show that it outperforms other feature-based approaches in predicting projects' daily collected funds by about 32-45%. This result indicates the learned competitiveness well reflect projects' daily performance. We also investigate the correlation between project characteristics and its final pledging result, in cases of different competitiveness levels. We find that high competitive projects are impressive with high pledging goals and high price rewards, comparing to medium and low competitive projects.

We further leverage the learned DMC model on two applications: predict the project success, and estimate days taken to achieve pledging goal. The experimental results show that with the DMC model, by simply adding competitiveness learned by DMC to existing feature-based approaches can achieve 85% in accuracy when predicting project success, and less than 7 days error when predicting number of days to hit the pledging goal.

For future works, we plan to expand our scope of studying to different market scenarios in order to investigate the potentially different competition phenomenon across markets. We also plan to explore different models that can capture the market competition phenomenon.

## ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation grant No.: 1717084.

## REFERENCES

- [1] J. An, D. Quercia, and J. Crowcroft. Recommending investors for crowdfunding projects. In *WWW*, pages 261–270. ACM, 2014.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [3] J. Boone. Competitive pressure: the effects on investments in product and process innovation. *The RAND Journal of Economics*, pages 549–569, 2000.
- [4] J. Chung and K. Lee. 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*, pages 211–220. ACM, 2015.
- [5] N. Desai, R. Gupta, and K. Truong. Plead or pitch? the role of language in kickstarter project success.
- [6] M. D. Greenberg, B. Pardo, K. Hariharan, and E. Gerber. Crowdfunding support tools: predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1815–1820. ACM, 2013.
- [7] W. Jank, G. Shmueli, M. Dass, I. Yahav, and S. Zhang. Statistical challenges in ecommerce: Modeling dynamic and networked data. *INFORMS Tutorials in Operations Research*, pages 31–54, 2008.
- [8] J. Kim, M. Lee, D. Cho, and B. Lee. An empirical analysis of semantic network in online crowdfunding: evidence from kickstarter. In *Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart connected World*, page 44. ACM, 2016.
- [9] Y. Lin, P. Yin, and W.-C. Lee. Modeling menu bundle designs of crowdfunding projects. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management*. ACM, 2017.
- [10] C.-T. Lu, S. Xie, X. Kong, and P. S. Yu. Inferring the impacts of social media on crowdfunding. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 573–582. ACM, 2014.
- [11] L. Makowski and J. M. Ostroy. Perfect competition and the creativity of the market. *Journal of Economic Literature*, pages 479–535, 2001.
- [12] L. Mia and B. Clarke. Market competition, management accounting systems and business unit performance. *Management Accounting Research*, 10(2):137–158, 1999.
- [13] T. Mitra and E. Gilbert. 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*, pages 49–61. ACM, 2014.
- [14] R. Ollman and M. Billington. The deadline model for simple reaction times. *Cognitive Psychology*, 3(2):311–336, 1972.
- [15] J. M. Podolny. A status-based model of market competition. *American journal of sociology*, pages 829–872, 1993.
- [16] M. E. Porter. *Competitive strategy: Techniques for analyzing industries and competitors*. Simon and Schuster, 2008.
- [17] V. Rakesh, J. Choo, and C. K. Reddy. Project recommendation using heterogeneous traits in crowdfunding. In *Ninth International AAAI Conference on Web and Social Media*, 2015.
- [18] V. Rakesh, W.-C. Lee, and C. K. Reddy. Probabilistic group recommendation model for crowdfunding domains. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 257–266. ACM, 2016.
- [19] J. Reiman and P. Leighton. *The rich get richer and the poor get prison: Ideology, class, and criminal justice*. Routledge, 2015.
- [20] P. W. Roberts. Product innovation, product-market competition and persistent profitability in the us pharmaceutical industry. *Strategic management journal*, 20(7):655–670, 1999.
- [21] K. M. Schmidt. Managerial incentives and product market competition. *The Review of Economic Studies*, 64(2):191–213, 1997.
- [22] J. Solomon, W. Ma, and R. Wash. Don't wait!: How timing affects coordination of crowdfunding donations. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 547–556. ACM, 2015.
- [23] G. J. Stigler. Perfect competition, historically contemplated. *Journal of political economy*, 65(1):1–17, 1957.
- [24] X. Vives. Innovation and competitive pressure. *The Journal of Industrial Economics*, 56(3):419–469, 2008.
- [25] Wikipedia. Competition (economics) — wikipedia, the free encyclopedia, 2017. [Online; accessed 3-January-2017].
- [26] Wikipedia. Kickstarter — wikipedia, the free encyclopedia, 2017. [Online; accessed 3-August-2017].
- [27] A. Xu, X. Yang, H. Rao, W.-T. Fu, S.-W. Huang, and B. P. Bailey. Show me the money!: an analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 591–600. ACM, 2014.
- [28] H. Zhang, G. Kim, and E. P. Xing. Dynamic topic modeling for monitoring market competition from online text and image data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1425–1434. ACM, 2015.
- [29] D. Zvilichovsky, Y. Inbar, and O. Barzilay. Playing both sides of the market: Success and reciprocity on crowdfunding platforms. Available at SSRN 2304101, 2015.