Inferring the Impacts of Social Media on Crowdfunding

Chun-Ta Lu, Sihong Xie, Xiangnan Kong, and Philip S. Yu
University of Illinois at Chicago, Chicago, IL 60607, USA
{clu29, sxie6, xkong4}@uic.edu, psyu@cs.uic.edu

ABSTRACT

Crowdfunding – in which people can raise funds through collaborative contributions of general public (i.e., crowd) – has emerged as a billion dollars business for supporting more than one million ventures. However, very few research works have examined the process of crowdfunding. In particular, none has studied how social networks help crowdfunding projects to succeed. To gain insights into the effects of social networks in crowdfunding, we analyze the hidden connections between the fundraising results of projects on crowdfunding websites and the corresponding promotion campaigns in social media. Our analysis considers the dynamics of crowdfunding from two aspects: how fundraising activities and promotional activities on social media simultaneously evolve over time, and how the promotion campaigns influence the final outcomes. From our investigation, we identify a number of important principles that provide a useful guide for devising effective campaigns. For example, we observe temporal distribution of customer interest, strong correlations between a crowdfunding project's early promotional activities and the final outcomes, and the importance of concurrent promotion from multiple sources. We then show that these discoveries can help predict several important quantities, including overall popularity and the success rate of the project. Finally, we show how to use these discoveries to help design crowdfunding sites.

Categories and Subject Descriptors

H.3.4 [**Information Systems Applications-Systems and Software**]: Information networks; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Experimentation

## Keywords

Crowdfunding, social media, early prediction

1. INTRODUCTION

Crowdfunding, or crowd-sourced fundraising, provides a revolutionary way to support ideas and projects across a number of industries, such as technology, music, film and art. For instance, Kickstarter is one of the largest websites for crowdfunding. Three of the movie projects funded by crowd investors in Kickstarter were nominated for Oscars 2013, and one of them won the Oscar Award in the short documentary category. A recent report [1] indicates that global crowdfunding raised nearly $2.7 billion for more than 1 million campaigns in 2012, and is expected to grow into a $5.1 billion industry by 2013. Although crowdfunding has been drawing substantial attention from the business sector, many aspects of crowdfunding have not received rigorous examination [15] in research communities. In particular, social influence, the key to the success of crowdfunding projects, has not been investigated thoroughly before in data mining communities.

Unlike traditional ways of fundraising, where the enterprises used to receive a small number of major investments from venture capital firms, in crowdfunding, the enterprise receives support through a large number of small contributions from customers. As a consequence, crowdfunding presents a mixture of entrepreneurship with social network participation, in which the customers play an unexpected role: investors. Different to conventional investors in venture capital firms, customers who have invested a project may also be likely to promote the project among his/her friends through online social networks. As a fast and far reaching channel to broadcasting information, social media (e.g., Twitter, Facebook, and GooglePlus) plays an important role in promoting crowdfunding projects. Both enterprises and customers can keep track of a crowdfunding campaign on social media, and obtain useful insights in real-time. For example, enterprises can analyze social media data to forecast the prevalence of the initiatives in order to adjust strategies accordingly. While customers can also make use of social media to identify promising projects that are more likely to be successful and lead to more rewards for the investments.

Predicting the outcomes of crowdfunding projects is very important for both enterprises and customers. Previous research on social media mainly focuses on modeling how an event spreads out in social media to predict the popularity of the event. Examples include modeling patterns of popularity evolution [3, 20], analyzing social dynamics[12], and predicting speed and scale of social influence [21, 11]. However, to the best of our knowledge, none of the existing works studied how a crowdfunding campaign develops
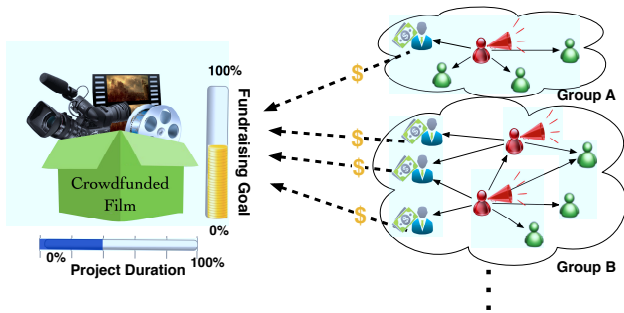
Figure 1: Example of the process of crowdfunding, where the red persons are the promotors, the blue persons are the backers, while the green persons are just potential targets.

in social media or what effects does social media have on the crowdfunding projects.

Crowdfunding has several unique properties that make it different from conventional information diffusion process in previous works. First, crowdfunding is usually a time-constrained process which is quite different from the ever-lasting process of information diffusion. For example, the fundraising process of a Kickstarter project can last from one day to 60 days[1]. Second, a project in crowdfunding should set a certain goal for fundraising, while the conventional information diffusion process doesn't have a collective ob-jective/goal from the crowd. Third, crowdfunding involves a variety of roles, including the promoters who disseminate information about the project in social networks and the backers who pledge funds on the project. Due to these unique properties, the mechanism of crowdfunding cannot be assessed by prior works [3, 20, 12, 21, 11] effectively.

Figure 1 illustrates the process of crowdfunding. In Figure 1, the left part is a crowdfunding project, which has a pre-specified project duration and a fundraising goal to be achieved. The right part is the campaign of the project on social media, which consists of multiple groups of partici-pants in the process of social promotion. In group A, for instance, there is a promoter who informs his/her friends of the project and one of them is influenced to invest in the project. The dashed lines in the middle of Figure 1 show the investments from backers affected by social networks. The key objective of this work is to infer the impacts of so-cial networks on the fundraising results of a crowdfunding project, such as the number of backer of the project and the successfulness of the fundraising goal.

To infer the social impacts, we study two concrete tasks motivated by the above potential applications for both enter-prises and customers. The first task is the prediction of the number of backers: given the social activities within social media during the early stage of a project, can we predict how many customers will contribute towards the project? The second task is the prediction of whether a venture will receive sufficient funding at the end of crowdfunding pro-

---

[1]Statistically, projects lasting more than 30 days have lower success rates to achieve its goal, since the longer dura-tion incites less urgency, encourages procrastination, and tends to fizzle out. `http://www.kickstarter.com/blog/shortening-the-maximum-project-length`

cess. Specially, given the information of a project, the funds collected so far and the online social media activities of the project, can we predict how likely its fundraising goal will be reached?

In order to address these two tasks, we begin by investi-gating the dynamics of crowdfunding from the fundraising results of crowdfunding projects as well as the correspond-ing campaigns on social media. From our investigation, we identify some important principles that not only help infer the impacts of social media but provide a useful guide for de-vising effective campaigns. The first principle we identify is that the temporal distribution of customer interests regard to a project is reciprocally affected by both the freshness and the remaining duration of the project. The launch of the project sparks a burst of interests in which the majority of customers and funds arrive. The customer interests then decay significantly until the approach of the project deadline again arouses customers to achieve the fundraising goal of the project. Therefore, in order to bring in more contribu-tions before customer interests decay, it is critical to identify potential bottlenecks so as to adjust campaign strategies by predicting the eventual outcomes at the very beginning.

Second, we observe that the results of a project are much more correlated to the early promotional activities on social media rather than its own properties (e.g. project dura-tion and fundraising goal). More precisely, we find that the number of backers of a project has strong correlation to the volume of its promotional activities, while its success rate is more correlated to the design of its promotion campaign. In other words, a project is popularized by massive promo-tion, whereas the foundation of its success is established in the intensive interactions between participants. This under-standing reinforces our premise that promotional activities in social media have positive impacts on the crowdfunding projects, and further hints the critical features for the two tasks we have defined.

Last but not least, we notice that the crowdfunding cam-paign of a project is driven by both the social promotion in social media and external stimulations outside social media. One can imagine that a campaign is composed of several groups of participants in social media. The social promo-tion from promoters to their friends expand the size of each group. In the mean time, the external stimulations from different channels expose the project to more groups. Thus, in order to spread a project to more contributors, the bet-ter way to process the promotion is to stimulate from social media as well as from other channels.

Overall, our goal is to contribute to a broader investiga-tion of this perspective on crowdfunding. Our performance on the two tasks suggests that features arising from the cam-paign activity in social media such as Twitter can provide important information beyond simply considering the data from crowdfunding sites.

2. BACKGROUND

Crowdfunding originated from the concept of crowdsourc-ing, in which a crowd of people collaboratively contribute to the solution to a task. In crowdfunding, participants play the roles of promoters and investors in support of proposed initiatives. Therefore crowdfunding presents a mixture of entrepreneurship and social network participation.

In crowdfunding, a proposer launches a project and re-quests certain amount of money that needs to be raised in

a specified duration [15, 16]. If they are interested in the project, investors can contribute funds to it. If the fundraising goal is reached, the funds are awarded to the proposer, and equal benefits are committed to the investors. Most of the crowdfunding sites (e.g. Kickstarter) further apply an all-or-nothing principle: if the fundraising goal is not reached by the deadline, all the funds are returned to the investors, while all commitments are cancelled. Such an all-or-nothing principle helps investors to avoid the risk of paying for nothing.

Besides fundraising activities on crowdfunding websites, typically there are also fundraising-related social activities on social networks like Twitter. Specifically, novel and creative crowdfunding projects are notably discussed on social networks where users are willing to share their interests, provide or respond to comments, promote and announce news of the projects. These activities on social networks can be seen as a complementary data source to crowdfunding websites and provide more insights into the crowdfunding process. We focus on a crowdfunding website called Kickstarter.com and related tweets on Twitter, since they exhibit the basic elements we wish to study, namely, the co-occurring fundraising and social activities on Tweeter. Our observations and methodologies are applicable to other crowdfunding hosts, so long as they allow promotions on social media.

3. DATASET CHARACTERISTICS

We start our presentation by describing the data acquisition and processing for our analysis. We then provide the statistical characteristics of our dataset.

3.1 Data acquisition and processing

Crowdfunding data: From Kickstarter, we obtained all the projects launched after Nov. 2012 and completed before Apr. 2013. For each project we recorded its proposer, duration, webpage URL, shorten URL[2], fundraising goal and available pledging options. Within each project's duration we recorded its total pledged amount and total number of backers on a daily basis. We noticed that there were some trial projects having no backers or extremely low fundraising goals (less than $100), and we removed them from our dataset since none of them reached their fundraising goals.

Twitter data: We gathered all the tweets regarding Kickstarter from Nov. 2012 to Apr. 2013. The tweets are extracted by the Twitter stream API with the keywords "kickstarter" and "kck.st/"[2], ensuring they contain timestamps, authors, mentioned users, tweet texts, and URLs for our analysis. For each tweet's author, we queried Twitter API for the metadata about the author as well as the followers and followees of the author. Since a tweet is limited to 140 characters, most URLs are shortened by a URL shortening service. We expanded them to get the original URLs.

Data preprocessing: A crowdfunding project could be shared on different channels and referred to by various titles. An effective promotion of a project should mention its official URL to encourage others to visit it. For each project we consider only the tweets mentioning its URL within its project duration. We call the author of such a tweet a promoter. Besides, Kickstarter provides a Twitter shortcut for users to share the news of sponsoring a project in a fixed for-
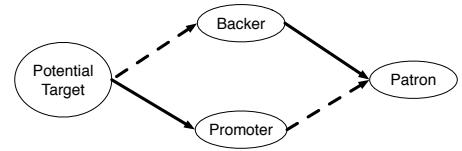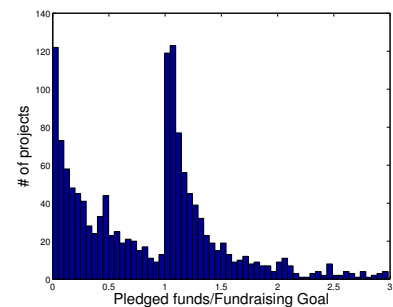

Figure 2: Transitions of roles in crowdfunding. Solid lines show the action of broadcasting and dash lines show the action of investment.

# of projects	1,521
# of successful projects	841
# of backers	145,032
# of tweets	62,473
# of promoters	39,051
# of patrons	16,666
# of mentioned users	14,415
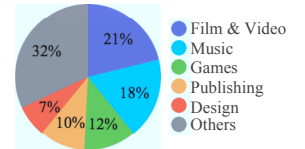
Table 1: Statistics of the dataset

**Figure 3: Percentage of project categories.**

Figure 4: Distribution of the ratio of total pledged amount to the fundraising goal. An additional 120 projects (7.8% of all projects) were funded more than 3x.

mat (e.g. "I just backed project_title @kickstarter"). Hence, we further filter tweets with the keyword "backed" and the project title, and call the author of such a tweet a patron. In this paper, we distinguish the patron in Twitter from the backer in Kickstarter by assuming that a patron is not only a backer who puts a pledge on a Kickstarter project but also tweets the news of sponsoring it. Figure 2 illustrates the transition of roles in crowdfunding. In Figure 2, a potential target would become a backer of a project if he/she invests in the project, or become a promoter if he/she broadcasts the information of the project. If he/she both invests and promote the project, he/she becomes a patron.

We further filter out projects that were seldom discussed (less than 5 tweets) in the Twitter dataset, simply because there are not enough data to correctly evaluate the crowdfunding campaign through Twitter. Table 1 and Figure 3 show the basic statistics of our dataset after filtering.

3.2 Statistical characteristics of dataset

The distribution of the ratio of total amount of pledged funds to the fundraising goal can be found in Figure 4. It illustrates the fact that most Kickstarter projects either fail significantly or meet their goals by relatively small margins. Among all the failed projects, only twenty-five percent of
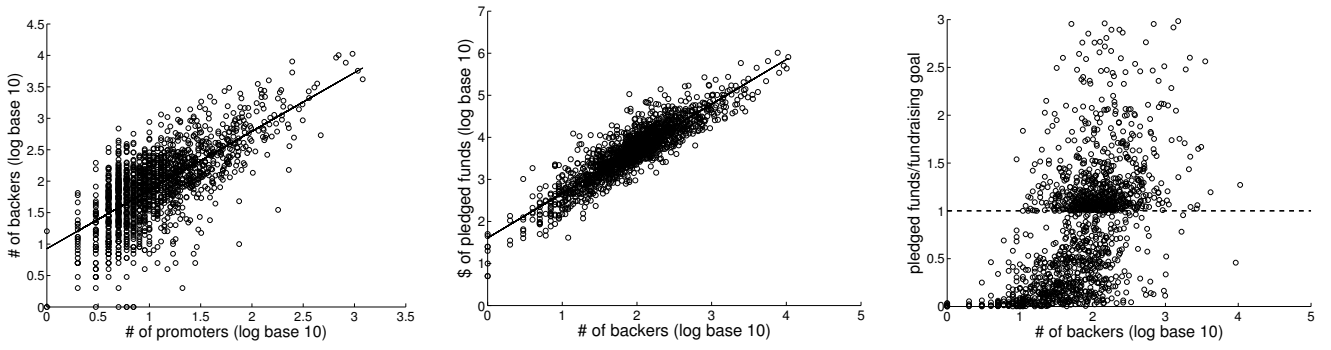
(a) Number of backers versus number of promoters

(b) Amount of pledged funds versus number of backers

(c) Ratio of pledged funds to fundraising goal versus number of backers

Figure 6: Each figure shows a scatter plot of all the projects. (a) Number of backers plotted against the number of promoters, where the correlation coefficient is 0.79. (b) Amount of pledged funds plotted against the number of backers, where the correlation coefficient is 0.9. (c) Ratio of pledged funds to fundraising goal plotted against the number of backers. 44% of projects below the dash line, y=1, which means they failed to reach their fundraising goals.
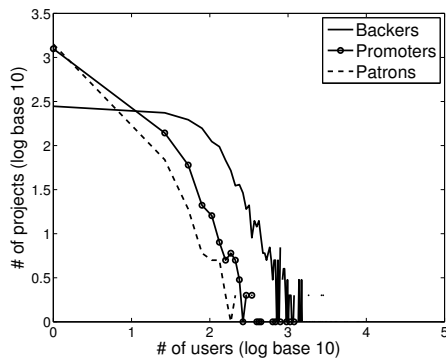
Figure 5: Log distribution of users and projects.

projects raise 50% of their goals. Fifty percent of projects that succeed are funded 25% or less over their goals, and only 7.8% of projects received over three times their goals. The margins of failure and success might be the result of the all-or-nothing policy. When a project shows a good chance of success, participants are more likely to invest to help the project to reach the goal.

Figure 5 displays the distribution of different types of users (i.e. backers, promoters and patrons) and the corresponding number of projects. In general, the distribution of social media users (promoters and patrons) follows the power law [3], and very few projects (1.5% of projects) are discussed by more than 100 social media users. However, the distribution of backers is unlike the distribution of social media users. It shows that there are fewer projects with few backers than one would expect from the power law, especially for the projects with less than 100 backers. The unexpected base of support seems owing to the contribution from the project

proposers and their friends. Thus, simply using the number of promoters might not predict the number of backers well.

In spite of that, there are non-negligible connections between the activities in social media and the results of crowdfunding projects. Consider the number of backers and the number of promoters in Figure 6(a) as an example. We can observe a strong correlation (0.79) between these two variables in log scale. This relationship provides a hint that social media activities are highly correlated with fundraising activities. In Section 5, we will further analyze how social activities affect the outcomes of crowdfunding projects.

4. PREDICTION TASKS

In this section we introduce two prediction tasks, which are based on observations of phenomena occur naturally in crowdfunding: the first is predicting the number of backers of a project and the second is predicting whether a project will receive sufficient funding. Our objective in formulating these tasks is to use them as an analytic framework, which is of practical interest to proposers and investors of crowdfunding projects.

In both prediction tasks, we focus on the early predictions that use only the data in the very early stages of crowdfunding projects to predict the outcomes of the projects. As we will see in Section 5.1, customer interests regard to a project decay significantly as the project evolves. In order to receive sufficient funds to support a project, it is important to identify potential bottlenecks as early as possible and adjust current strategies to attract more contributions before customer interests vanish. Thus, early prediction of the outcomes of projects, as well as understanding the associated properties and signals, are critical to the effectiveness of crowdfunding campaigns.

Predicting the number of backers. A crowdfunding project needs to raise funds and only the enrolled backers make actual monetary support. We can view the backers of a project as the buyers of a commercial product. Moreover, there is an extremely high correlation (0.9) between the amount of raised funds and the number of backers (both

---

[3] Although all the projects in our dataset are referred by at least 5 tweets, some of them have less than 5 promoters because a promoter may tweet more than once.

in log scales), as shown in Figure 6(b). This high correlation suggests that the amount of raised funds strongly depends on the number of backers. Therefore, early prediction of the number of backers would help us estimate the amount of funds a project may receive.

Predicting whether a project will reach its fundraising goal. Although the funds of projects are highly correlated to the number of their backers, different projects may have dramatically different fundraising goals. Projects with a large amount of backers may not reach their fundraising goals. For example, Figure 6(c) shows that certain projects with more than 1,000 backers can fail to reach their fundraising goals (as shown by the points below the dash line, y=1). Our second task forms a natural complement to the first: other than predicting the number of backers of a project, we try to predict whether a project will receive sufficient funds to reach its fundraising goal or not.

Early prediction of the success rate of projects is clearly useful to investors as well as to crowdfunding services. Investors can reduce the waste of time and money by turning their attention to the projects that show more promising signs of success. The quality of crowdfunding platform can also be improved by extending current recommendation services to take into account the future success rate.

With these two prediction tasks in mind, we explore various features that affects the project results. After this exploration, we will show that the information we derive from social media helps us accurately predict both properties in Section 6.

5. DYNAMICS OF CROWDFUNDING

In this section we investigate the dynamics of crowdfunding from the processes of fundraising activities in crowdfunding sites as well as promotional activities in social media. We group our analysis into two parts to answer the following two questions: how participants response to a crowdfunding project over time, and how the campaigns in social media influence the eventual outcomes. In these two parts, we identify some basic principles that will be useful when we develop techniques for our prediction tasks.

5.1 The phases of crowdfunding activities

We start by investigating the process of crowdfunding activities along the span of a project's lifetime. Prior empirical study [7] has shown that the attention of online sharing-content peaks at the beginning and significantly decays after that. From this study one would expect that a burst followed by a power-law distribution governing the intensity of crowdfunding activities.

However, unlike conventional online content, a project has a specific deadline, after which no more action can change the results of the project. This policy arouses the potential contributors to put money into the project when the deadline is approaching. This is as if the remaining amount of activities in the long-tail of the power-law after the deadline would aggregate to the period before deadline.

Following this observation, we expect that there are threes phases in a project's lifetime: (1) starting bursty phase, when it is launched in the crowdfunding sites, (2) stationary phase, after the starting bursty follows a long period of relatively little activities and (3) final bursty phase, when its deadline is approaching.
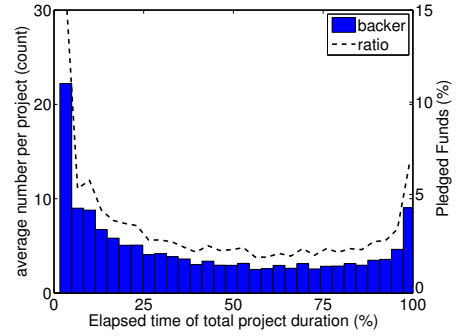

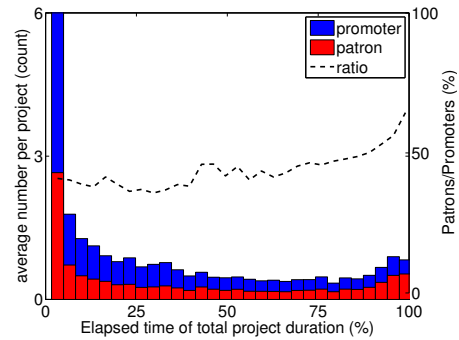Figure 7: Histogram of the fundraising activities in Kickstarter.

Figure 8: Histogram of the promotional activities in Twitter.

Phases of fundraising activities. In Figure 7, we examine the process of fundraising activities in Kickstarter. The blue bars represent the number of backers and the dash line shows the fraction of the raised funds in each period to the total raised funds at the target deadline. We find that the starting bursty phase happens in the first quarter of project duration, the final bursty phase appears in the last ten percent of project duration, and between these two phases is the stationary phase. A considerable portion of contribution a project receives can be attributed to the starting bursty phase, in which half of its backers invest in it and they contribute around 50% of its total raised funds. Although the majority of funding is acquired, only 14% of projects can reach their fundraising goals within the starting bursty phase. Around 60% of projects depend on the funds received in the final bursty phase to determine whether they can succeed or not (25% of projects among them succeed).

Phases of promotional activities. We next assess the process of promotional activities on Twitter, as shown in Figure 8. The blue bars represent the number of promoters, the red bars represent the number of patrons, and the dash line expresses the ratio of patrons to promoters in each period. We observe there also exist three similar phases in Figure 8 as in Figure 7: major amount of promotional activities in the first quarter, little attention in the middle period, and raising interests in the last ten percent of project duration.

Moreover, we notice that the patron/promoter ratio has an upward trend. It implies that the priority of crowdfund-

(a) Number of backers

(b) Fraction of pledged funds to fundraising goal
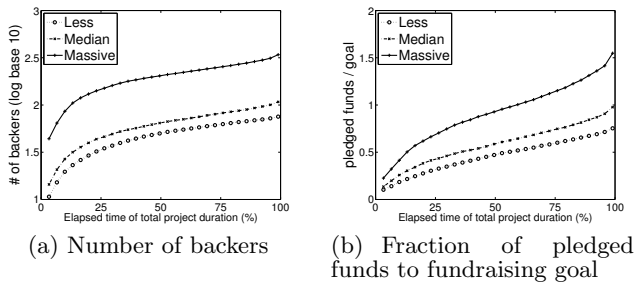
Figure 9: The comparison of fundraising results for different levels of promotion in starting bursty phase.

ing activities is shifting from distributing the information to sharing the news of backing when approaching the deadline. We believe this shift is the result of the fact that it takes certain amount of time for a project to convince others to invest in, and the approaching deadline expedites the decision making in investments. Hence the longer it has been the higher percentage of the crowdfunding activities will shift from promoting to backing.

However, due to the decay of interests as time goes by, a project may not benefit much from this activity shifting. Hence, the prediction tasks we target to solve will focus on predicting the results from the data available within the starting bursty phase, as they will be more useful in practice.

Effects of promotional activities. The similar phases and activity shifting arise a series of interesting questions: Does this mean the fundraising activities are correlated to the promotional activities? Could we expect the projects with higher promotional activities will have better fundraising results? Could we forecast the results of crowdfunding projects from the early promotional activities?

We answer these questions by comparing with different levels of promotional activities in starting bursty phase, which is measured as the number of promoters accumulated by 25% of project duration. A project with the number of promoters between 1 and 5 is called a "less discussed project". For projects with 6-10 (11-100, respectively) promoters, we call them "median (massively, respectively) discussed projects". In Figure 9, we find that the projects with more promoters in starting bursty phase are always better in the number of backers and also in the ratio of pledged funds to fundraising goal. We stress that we are not claiming a direct causal relationship between promotional activities and the fundraising results; rather we are pointing out that there exist strong correlations between the fundraising results and early promotional activity. We will further discuss them in the following section.

5.2 Correlations between social promotion and fundraising results

In the previous section we observed the phases of crowdfunding activities, and demonstrated the effects of early promotional activities. Now we turn our attention to the underlying factors of crowdfunding campaigns on social media that lead to the difference in their prevalence. Specifically, we analyze correlations between the fundraising results of

projects and the features extracted from corresponding campaigns on social media.

Here we treat the campaign of a crowdfunding project on a social network as the process of an information propagation on the network. A node in the campaign graph is a promoter, which is an author of a tweet regarding the project. A directed edge between a pair of nodes, e.g. $e_{u,v}$, expresses that the information of the project flows from one promoter u to the other promoter v. There are two criteria for this information flow: 1) the earliest tweet from promoter v regarding to the project is posted after that from promoter u, and 2) promoter v follows promoter u or was mentioned in promoter u's tweet. If a promoter has no inward edges, there is no information flows to the promoter and the promoter must be motivated by external stimulations outside the monitored social network. Without loss of generality, the process of crowdfunding campaigns can be traced by the development of the information propagation.

After constructing the graphs of crowdfunding campaigns, we extract features from the structure of the graphs as well as from the crowdfunding activities. Overall, the features are explored from three different aspects: (1) properties of each project in crowdfunding sites, (2) activities of participants in crowdfunding campaigns, and (3) structures of the campaign graphs. The full set of features we consider is as follows:

Project features (S_A), 7 features total: project duration, elapsed days since the launch, fundraising goal, median amount of pledge options, number of backers, ratio of current raised funds to fundraising goal, and average amount of pledge per backer.

Social activity features (S_B), 6 features total: number of tweets, number of promoters, number of patrons, number of unique mentioned users, fraction of promoters from external stimulations, and fraction of patrons from external stimulations.

Social structure features (S_C), 7 features total: average number of followers of promoters, median number of followers of promoters, number of edges, diameter (largest shortest distance), number of weakly connected components, number of triads, and global clustering coefficient [9] (average local clustering coefficient [10, 23] of all nodes).

Note that the last two features in S_B intend to capture the effect of external influence. The fraction of promoters (patrons) from external stimulations is measured as the fraction of the number of promoters (patrons) with no inward edges to the total number of promoters (patrons). In S_C, the edges represents the appearances of internal influence, and the diameter means the largest geodesic distance the cascade internal influence can reach from one node to the other one. The weakly connected components represents the groups in the crowdfunding campaigns. The number of triads and the global clustering coefficient are used to study the strength and density of the campaigns.

Correlations between proposed features and fundraising results. We now evaluate the correlations between the 20 features mentioned above and the final fundraising results at target deadline. The features are extracted from the information available in the starting bursty stage. Table 2 and Table 3 display the correlation values for the number of backers and the ratio of fundraising goal achievement at the target deadline, respectively. For example, in Table 2, for the row on the number of promoters, the entry on the column

Feature	project duration (%)		
	5	10	25
# of backers	0.82	0.89	0.92
# of patrons	0.79	0.81	0.85
# of promoters	0.77	0.78	0.79
# of tweets	0.73	0.70	0.65
# of weakly connected components	0.70	0.69	0.68

Table 2: Correlation of the total number of backers at target deadline and the features extracted at specified time.

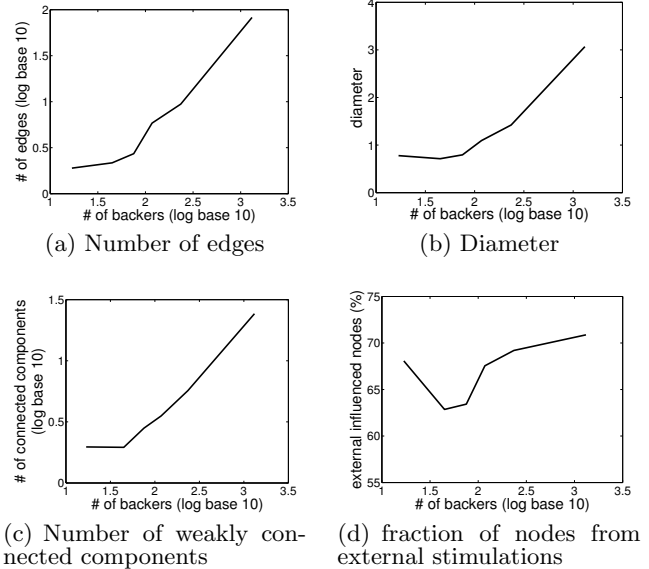| Feature	project duration (%)		
	5	10	25
raised funds/goal	0.77	0.89	0.95
# of triads	0.47	0.46	0.45
# of edges	0.43	0.43	0.38
# of patrons	0.26	0.26	0.24
# of backers	0.18	0.29	0.30

Table 3: Correlation of the ratio of raised funds to fundraising goal at target deadline and the features extracted at specified time.

of 5% means that the correlation between the accumulated number of promoters by the end of the 5% period of the fundraising duration to the number of backers at the target deadline is 0.77. This correlation increases slightly to 0.78 and 0.79 at the end of the 10% and 25% periods, respectively.

For each table, we only report the top 5 features that have the highest correlation at each time frame. As shown in both tables, the feature that has the best correlation is the fundraising result itself. This indicates the early status of the fundraising result has good predictive power on its final value, and we should consider it as baseline for each task. Interestingly, the other top features for both objectives are extracted from social media, instead of the properties of a project on its own.

Table 2 shows that the number of backers is strongly correlated to the volume of promotional activities, i.e. the number of patrons, the number of promoters, and the number of tweets. On the other hand, Table 3 shows that the ratio of fundraising goal achievement is more correlated to the structure of campaign graphs, i.e. the number of triads and the number of edges. These observations are due to the fundamentally different properties of the volume and network structures. The number of backers is measured as the quantity of support and is primarily affected by the volume of promotion, whereas its success is determined by the quality of support, i.e. the amount of money. A popular project may receive lots of very small contributions, e.g. $1 dollar, from general public but still not be able to reach the fundraising goal. However, a successful project would target specific groups that have interests in it and acquire more money from them. The number of edges and the number of triads both suggest the strength of information flows within groups. A large number of edges and/or triads show a sign of the intensive interests among groups, and thus they are more correlated to the ratio of fundraising goal achievement.

Concurrent processes of social promotions and external stimulations.

(a) Number of edges

(b) Diameter

(c) Number of weakly connected components

(d) fraction of nodes from external stimulations

Figure 10: Each figure shows, given the promotional activity in the starting bursty phase, the value of a particular property as a function of the number of backers at target deadline. Top row and bottom row display the properties from the aspect of social promotions and external stimulations, respectively.

Our above observations show that there are strong correlations between fundraising results and social features. We claim they are primarily due to the fact that if a project is more persuasive to intrigue authors in social networks to discuss it (either by social promotions or external stimulations), it is more attractive to investors. From the aspect of social promotions, we expect such a project to have a larger campaign graph. From the aspect of external stimulations, we expect there exists more groups in the campaign, each of which is represented by a weakly connected component as each component can be viewed as started by an external stimulus.

We consider the relationship between the number of backers and the social structure of its campaign in the starting bursty phase as an example to illustrate our claim. From the aspect of social promotions, as shown in Figure 10(a) and Figure 10(b), we notice that the number of backers increases with the number of edges and/or the diameter. Both features indicate the expansion of crowdfunding campaigns which is activated from the influences of social promotions. From the aspect of external stimulations, as can be seen in Figure 10(c), we observe the similar rise in the number of weakly connected components, which indicates the increase in the number of groups.

These results from both aspects suggest that social promotions and external stimulations concurrently activate the processes of crowdfunding campaigns. One may ask which one is more important? In Figure 10(d), we find that over 60% of nodes are from external stimulations. This reflects the fact that most authors access a project from external sources outside social media and motivated by personal interests to promote it. However, we also notice if there are

very few social promotions, which signals the unenthusiastic response in social media, the number of backers would be very low. Hence, the effective way to promote a project is to stimulate from social media as well as stimulate from other channels. From the features extracted from social media, we could identify the bottlenecks of current campaign and devise a better campaign strategy.

6. EXPERIMENTS

The previous section presented a set of properties that governing the process of crowdfunding. Now we show that this understanding is directly applicable to our two prediction tasks.

The practical applications of both tasks would be to predict project results as early as possible. Accordingly, we perform the following tasks using only the information available in the starting bursty phase (within 25% of project duration after the project is posted in crowdfunding sites). We consider the features presented in the previous sections for our prediction tasks. Note that besides focusing on prediction accuracy, we also aim at obtaining explanatory insights into what features are the most relevant and useful as signals for crowdfunding processes.

6.1 Predicting the number of backers

Recall from Section 4 that our first task is to predict the number of backers of a project at target deadline. As suggested in [20, 18], the Relative Squared Error (RSE) is more relevant and meaningful than absolute quadratic error for online content popularity prediction. We here adopt RSE to evaluate the performance of our first task. Let $N(p)$ be the total number of backers project p has received at target deadline, and $N(p|t)$ be the total number of backers predicted for project p based on data from the first t days. The RSE takes the form of

$$RSE = \left[\frac{N(p|t) - N(p)}{N(p)} \right]^2 = \left[\frac{N(p|t)}{N(p)} - 1 \right]^2 \quad (1)$$

For the collection of projects, denoted by C, the mean Relative Squared Error (mRSE) is defined as the arithmetic mean of the RSE values for all projects in C, that is:

$$mRSE = \frac{1}{|C|} \cdot \sum_{p \in C} \left[\frac{N(p|t)}{N(p)} - 1 \right]^2 \quad (2)$$

The correlation between crowdfunding activity and the future number of backers observed in previous section suggests that the number of backers can be estimated as a linear function of features. Thus, we apply a multivariate linear model [18] to predict the number of backers. We denote the feature vector of the project p at time t as $X(p,t)$. The future number of backers of the project can be estimated as

$$N(p|t) = \theta_t \cdot X(p,t) \quad (3)$$

where θ_t is the vector of model parameters and depends only on t.

Given a training set C, t, we can estimate the optimal values for the elements of θ_{t_i} as the ones that minimize the mRSE on C, i.e.:

$$\operatorname*{argmin}_{\theta_t} \frac{1}{|C|} \sum_{p \in C} \left[\theta_t \cdot \frac{X(p,t)}{N(p)} - 1 \right]^2 \quad (4)$$
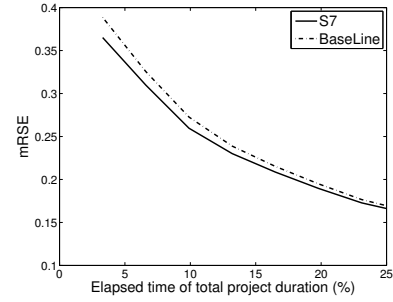
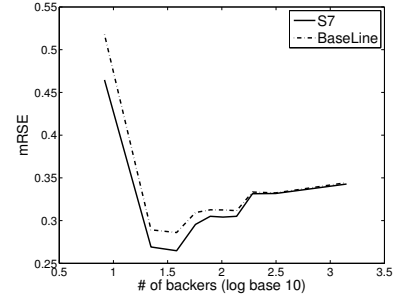**Figure 11: The performance of the prediction of the number of backers.**

**Figure 12: The mRSE shown as a function of the number of backers at target deadline.**

which is an ordinary least squares (OLS) problem that can be solved via a singular value decomposition [5, 18].

We conduct feature selection and find a core set of 7 features that we use in this task: number of backers, number of tweets, number of promoters, number of patrons, fraction of promoters from external influence, number of edges, and number of weakly connected components. Note that except the number of backers, all the other six features are extracted from social media. The first four in the six features are from social activity features, while the last two are from social structure features. We call this set S_7 and compare it against the baseline that only uses the number of backers.

We perform 5-fold cross validation and report the mRSE. Figure 11 gives the results of the prediction task. We can observe that the features extracted from social media (S_7) inspired by our previous analyses improve the performance of prediction. For instance, when using only the data available up to 5% of project duration, the mRSE produced by the S_7 and by the baseline are 0.337 and 0.357, respectively, leading to an error reduction of 5.5%. If extending the time window to 25% of project duration - after which half of backers have arrived (see Section 5.1) - both mRSE values can be significantly reduced to about 0.16. The performance gains from using social media features are gradually reduced. This is reasonable, as the scale of the backers reveals more information of its final count, predictions just become easier. However, it is less meaningful to do such predictions at this stage, as it might be too late for project owners to adjust their strategies.

To further understand the performance of prediction for different popularities, in Figure 12, we analyze the mRSE values produced by comparing S_7 with baseline at 5% of
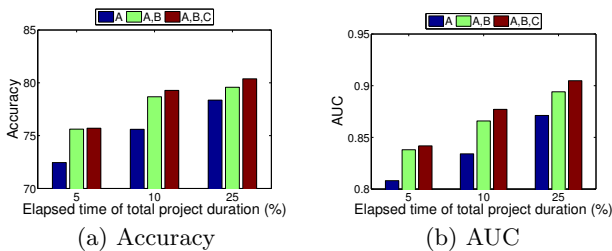
Figure 13: The performance of predicting the fundraising goal achievement.

project duration. Note that for the most challenging case when the number of backers is small or moderate, S_7 performs a lot better than the baseline as it can capture other clues from the social network attributes. For example, when the number of backer is 10, the mRSE produced by the S_7 and by the baseline are 0.46 and 0.52, respectively, leading to an error reduction of 9.6%. More importantly, the results show that the prediction made by using S_7 features outperform baseline for projects with any number of backers.

6.2 Predicting whether a project will receive sufficient funds

We further study the performance of the four feature classes on predicting if a project will achieve its fundraising goal. Specifically, we formulate this prediction task as a binary classification task. The response variable is whether the amount of pledged funds exceeds the fundraising goal.

After performing feature selection, we reduce our features to a smaller set of 8 features: 2 from S_A (number of backers, and ratio of current pledged funds to the fundraising goal), 2 from S_B (number of patrons and fraction of promoters from external influence), and 4 from S_C (number of edges, average number of followers, number of weakly connected component, and number of triads).

For interpretability and ease of comparison, we use a logistic regression classifier, perform 5-fold cross validation and report the classification accuracy and the area under the ROC curve (AUC). The results are reported in Figure 13. For example, in Figure 13(a), it shows that at 5% of the project duration, just using feature from the project property (S_A) as indicated by the label A, the prediction accuracy is around 0.725. Adding features from social activity (S_B) as indicated by the label (A,B) will increase the accuracy to 0.754 and further incorporating the social structure features (S_C) indicated by the label (A,B,C) will increase the accuracy to 0.757.

We notice that the properties of the crowdfunding projects on their own (S_A) have good predictive power. It gives 22% improvement in accuracy over random guessing. Although extending the time window can improve the prediction accuracy, it is less useful in practice. Moreover, extending the time window provides less improvement than adding social activity features (S_B), which gives a gain of 3-5% in accuracy and 3-4% in AUC in each time frame. Incorporating social structure features (S_C) in the prediction model further enhances the improvement. This improvement is clearer when the social structure has developed for a period of time (e.g. after 10% of project duration).

To fully appreciate the result, we emphasize that the results of 60% of projects are determined in the last ten percent of project duration. It is remarkable that we achieve 75% accuracy by using features available only up to five percent of project duration after the project was launched. From these results, we see that taking into account the rich information from social media improves performance of prediction on crowdfunding.

7. RELATED WORK

In this section, we will survey existing works about the determinants of the success of crowdfunding ventures. We then present some studies around the impact of social media.

Determinants of venture success: Crowdfunding has been drawing substantial attention from entrepreneurs, but relatively little notice from academics. Some studies focus on an important issue: the underlying factors of success and failure among crowdfunding ventures [6, 15]. To examine the determinants of venture success, Mollick et al. [15] conducted an analysis using logistic regression of the odds of successful funding from Kickstarter. Their result suggests that the size of a proposer's social network (e.g. the number of Facebook friends) help predict the success of crowdfunding efforts. However, there are only 16% of projects whose proposers have Twitter accounts in our dataset. In these 16% of projects, the number of followers of the project proposers in Twitter has very low correlations to the number of backers (-0.0117) and to the ratio of fundraising goal achievement (-0.0488). More importantly, the process of crowdfunding is not simply driven by the members in the proposer's social network; instead, it is more likely to rely on the information propagation in social media, as discussed in Section 5 and Section 6.

Information diffusion models: Social media platforms have recently emerged as valuable sources of realtime news, event updates, opinions, and advertising. Enormous works have been done on analyzing how an event spreads out in social media [14, 17, 22, 21, 2, 13, 4, 19]. For modeling the global influence of nodes, Yang et al. [22] proposed the Linear Influence Model that takes newly infected nodes as a function of the times when other nodes got infected in the past. However, the authors in [17] demonstrated that information reaches users not only through connections in their online social networks, but also through the influence external out-of-network sources, like the mainstream media. Thus, to solve our prediction tasks, we should consider internal influence from neighbors as well as the external influence outside the given social media.

Prediction of future popularity: There is a stream of research that has examined the prediction of the future popularity of online content. For example, the study in [12] presents stochastic models of user behavior on content-sharing sites (e.g. Digg[4]) for predicting popularity based on early user reactions to new content. Henrique et al. also applied historical information of early popularity measures to predict the future popularity of YouTube videos [18]. Despite that, crowdfunding campaigns has several unique properties, as discussed in the introduction, that make it different to conventional online-sharing content. Hence these approaches cannot be directly applied to our prediction tasks.

---

[4]http:// digg.com

Social media as word-of-mouth wisdom: Jansen et al. [8] used Twitter to investigate the structure of the postings, the types of expressions, and the movement in sentiments. However, the authors do not perform any analysis on the predictive aspect of Twitter. Asur et al. [4] demonstrated how Twitter can be used to predict real-world outcomes. They forecasted box-office revenues for movies by a linear regression model and further improved the result by classifying sentiments of tweets. The authors in [19] constructed time-constrained graphs from stock-market related tweets and estimated the correlation between stock-market events and micro-blogging data. These works show the possibility of using social media to predict future outcomes. Nevertheless, none of them investigate how social media create impacts to time-constrained events, such as electoral campaigns or crowdfunding campaigns.

8. CONCLUSION

As crowdfunding flourishes as a feasible fundraising strategy – through the collective efforts by participants who network and pool their money together – in many industries, it becomes useful to think about how to attract more contributions. In this paper, we presented the underlying connections between promotion campaigns in social media and the fundraising results of crowdfunding projects, and identified a number of important principles that guide the direction of devising an effective campaign. We also demonstrated that the features arising from social media can help improve the accuracy of predicting eventual outcomes, specifically how popular a project will be, and which projects are likely to successfully reach the fundraising goals.

In summary, our results provide a promising step towards inferring the impacts of social media on crowdfunding. We anticipate that further analysis could quantify personal interests of participants from historical behaviors, develop richer ways of assessing social influences among users, and suggest potential customers from collaborative filtering of personal interests and social influences.

Acknowledgment

9. REFERENCES

[1] Crowdfunding industry report: Market trends, composition and crowdfunding platforms. `http://www.crowdsourcing.org/editorial/2013cf-the-crowdfunding-industry-report/25107`.

[2] C. C. Aggarwal, S. Lin, and P. S. Yu. On influential node discovery in dynamic social networks. In *SDM*, pages 636–647, 2012.

[3] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini. A peek into the future: predicting the evolution of popularity in user generated content. In *WSDM*, pages 607–616, 2013.

[4] S. Asur and B. A. Huberman. Predicting the future with social media. http://arxiv.org/abs/1003.5699, 2010.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[6] G. Burtch, A. Ghose, and S. Wattal. An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets. *Information Systems Research*, 2013.

[7] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[8] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.

[9] A. Kemper. *Valuation of network effects in software markets. A complex networks approach.* Contributions to Management Science., 2010.

[10] G. Kossinets and D. J. Watts. Empirical Analysis of an Evolving Social Network. *Science*, 311(5757):88–90, Jan. 2006.

[11] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *ICWSM*, 2010.

[12] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *WWW*, pages 621–630, 2010.

[13] C.-T. Li, H.-P. Hsieh, S.-D. Lin, and M.-K. Shan. Finding influential seed successors in social networks. In *WWW*, pages 557–558, 2012.

[14] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.

[15] Mollick and E. R. The dynamics of crowdfunding: Determinants of success and failure. *Social Science Research Network*, 2013.

[16] M. J. Muller, W. Geyer, T. Soule, S. Daniels, and L.-T. Cheng. Crowdfunding inside the enterprise: employee-initiatives for innovation and collaboration. In *CHI*, pages 503–512, 2013.

[17] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *KDD*, pages 33–41, 2012.

[18] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *WSDM*, pages 365–374, 2013.

[19] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes. Correlating financial time series with micro-blogging activity. In *WSDM*, pages 513–522, 2012.

[20] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, Aug. 2010.

[21] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. In *ICWSM*, 2010.

[22] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *ICDM*, pages 599–608, 2010.

[23] Y. Zhao, G. Wang, P. S. Yu, S. Liu, and S. Zhang. Inferring social roles and statuses in social networks. In *KDD*, 2013.