

# How Medical Crowdfunding Helps People? A Large-scale Case Study on Waterdrop Fundraising

Junjie Huang<sup>1,2</sup>, Huawei Shen<sup>1</sup>, Qi Cao<sup>1,2</sup>, Li Cai<sup>3</sup>, Xueqi Cheng<sup>1</sup>

<sup>1</sup>CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> Beijing Zongqing Xiangqian Science and Technology Co Ltd, Beijing, China  
{huangjunjie17s, shenhuawei, caoqi, cxq}@ict.ac.cn, caili@shuidihuzhu.com

## Abstract

While online medical crowdfunding achieved tremendous success, quantitative study about whether and how medical crowdfunding helps people remains little explored. In this paper, we empirically study how online medical crowdfunding helps people using more than 27,000 fundraising cases in Waterdrop Fundraising, one of the most popular online medical crowdfunding platforms in China. We find that the amount of money obtained by fundraisers is broadly distributed, i.e., a majority of lowly donated cases coexist with a handful of very successful cases. We further investigate the factors that potentially correlate with the success of medical fundraising cases. Profile information of fundraising cases, e.g., geographic information of fundraisers, affects the donated amounts, since detailed description may increase the credibility of a fundraising case. One prominent finding lies in the effect of social network on the success of fundraising cases: the spread of fundraising information along social network is a key factor of fundraising success, and the social capital of fundraisers play an important role in fundraising. Finally, we conduct prediction of donations using machine learning models, verifying the effect of potential factors on the success of medical crowdfunding. Altogether, this work presents a data-driven view of medical fundraising on the web and opens a door to understanding medical crowdfunding.

## Introduction

ArtistShare<sup>1</sup> launched in 2003 as the Internet's first fan-funding (referred to today as "crowdfunding") platform for creative artists. After that, lots of online crowdfunding platforms have emerged (e.g., kickstarter<sup>2</sup>) and play a role of intermediates to bridge the fundraisers and backers. Different from commercial crowdfunding, medical crowdfunding is a way of fundraising for medical bills. It has become a popular choice for people with unaffordable health needs, especially for low-income and middle-income certain families with limited social welfare. Online platforms for medical crowdfunding are rapidly emerging in many countries. Several popular online medical crowdfunding platforms in-

clude GoFundMe<sup>3</sup> in the USA and Waterdrop Fundraising<sup>4</sup> in China. They all achieved immense success in the medical crowdfunding. Despite the proliferation of online crowdfunding, little is known about whether and how medical crowdfunding helps people, especially lacking quantitative study. Compared with GoFundMe in the USA which allows people to raise money for other life events (e.g., celebrations and graduations), Waterdrop Fundraising is focused on medical crowdfunding. Moreover, Waterdrop Fundraising is conducted on social platforms in China (e.g., WeChat). Therefore, spreading relevant fundraising information in social networks is an important factor for the success of crowdfunding campaigns. The relationships among users of social platforms are relatively intimate social relationships (e.g., family, relatives, and friends). Based on these social platforms, Waterdrop Fundraising gives some guides for fundraisers and has developed a convenient online payment system, making donations and sharing easier. These differences between GoFundMe in the USA and Waterdrop Fundraising in China raise new research demand. In other words, for this kind of newly emerging medical crowdfunding platform, i.e., Waterdrop Fundraising, quantitative researches on them will help to explore the trends and characteristics of online medical crowdfunding.

We illustrate how medical crowdfunding works in Figure 1. For online medical crowdfunding, there are usually three key components: **a fundraiser**, **a platform** and **the social network of the fundraiser**. We divide a fundraising campaign into three parts: before information diffusion, during information and after information diffusion. Before information diffusion, the fundraiser first submits some information on the platform, and the platform reviews the authenticity of the case. After that, the fundraiser begins to spread his fundraising request to the social network (e.g., the WeChat Moment or WeChat Group). In addition to re-sharing and donating, people in the social network can also verify the fundraising cases to increase the credibility of the fundraising campaigns. This verified information usually includes verifying the authenticity and expressing the relationship with the fundraiser. For example, *This patient is my friend; he is recently suffering a severe disease. I hope ev-*

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup><https://www.artistshare.com/>

<sup>2</sup><https://www.kickstarter.com/>

<sup>3</sup><https://www.gofundme.com/>

<sup>4</sup><https://www.shuidichou.com/>

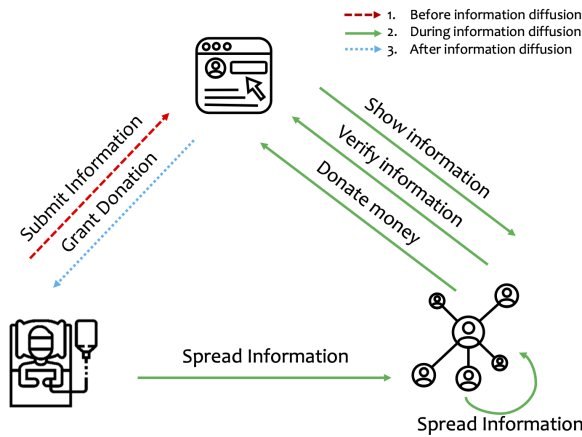


Figure 1: Medical fundraising on the web can be divided into three phases: (i) before the information diffusion, a fundraiser proposes a new case to the platform to conduct a preliminary review. (ii) during the information diffusion, users use their social networks to spread the crowdfunding information and call for donations. The platform collects verified information and donations from social networks. (iii) after the information diffusion, the platform grants the donations to the fundraiser.

*eryone gives him a hand.* People can easily continue these activities on social networks. The platform collects information, and donations from social networks and strengthens the crowdfunding campaigns. After the fundraising is completed or finished by the fundraiser, the platform will grant the collected money to fundraisers to treat the disease. The fundraising model illustrated in Figure 1 is adopted by most medical crowdfunding platforms in China.

Based on the forementioned fundraising model, we focus on answering two research questions in this paper:

1. Given the model in Figure 1, **what factors affect fundraising campaigns at different phases of information diffusion?** More specifically, before the information diffusion, what social status factors will affect the fundraising campaign? During the information diffusion, what social network factors will affect the fundraising?
2. Based on social status factors and social network factors, can we build a predictive model that predicts future fundraising result in the early stages of information diffusion?

**Organization** The rest of the paper is organized as follows. Firstly, we introduce the relevant background and related works. Then, we describe our dataset and do data cleaning. The dataset comes from a popular online fundraising platform in China, which spans a month in March of 2019, including a total of 27,000 cases. We find a low achieved ratio in these fundraising campaigns. Further, we quantitatively study the factors which affect the fundraising, including the impact of social status and social network. Last but not least, we formulate donations prediction tasks using different machine learning methods. We adopt some methods from popu-

larity prediction and achieve good predicted results. Finally, we offer our concluding thoughts and discuss future work.

## Background and Related Work

In this section, we will briefly introduce some backgrounds and related works about medical crowdfunding.

### Crowdfunding Mechanism

Crowdfunding is defined as an open call mostly through the Internet for the provision of financial resources by a group of individuals instead of professional parties either in the form of donations, in exchange for a future product, or exchange for some kind form of reward (Belleflamme, Lambert, and Schvienbacher 2014; Schvienbacher and Larralde 2010; Gerber and Hui 2013). Crowdfunding typically contains three participating stakeholders: *the project initiators* who seek funding for their projects, *the backers* who are willing to back a specific project, and *the matchmaking crowdfunding platforms* acting as intermediaries (Gierczak et al. 2016; Belleflamme, Lambert, and Schvienbacher 2014). In this paper, we refer to three participate stakeholders as **fundraisers**, **donors**, and **platforms**. Many researchers have discussed and studied the relationship between the three parts, their motivations, benefits, and risks (Gerber and Hui 2013; Haas, Blohm, and Leimeister 2014; Belleflamme, Lambert, and Schvienbacher 2014; Gierczak, Bretschneider, and Leimeister 2014). It is worth mentioning that past research has mainly discussed commercial crowdfunding, such as Kickstart. With the development of crowdfunding, non-profit organizations and government organizations have also used crowdfunding for social good (e.g., alleviate poverty<sup>5</sup>). Medical crowdfunding is different from the other models mentioned above. We compare these different crowdfunding models from platform, target, risk, and time in Table 1.

From Table 1, medical crowdfunding is primarily fundraising for treating personal diseases, which uses social networks to disseminate and verify the information. The utility of information diffusion is much higher than others. It can be considered that social capital (Lin, Cook, and Burt 2001) has played an important role in crowdfunding for serious illness. Besides, the platform effects (Malik and Pfeffer 2016) also show some impacts in medical crowdfunding. Based on this characteristic, we analyze the social factors that affect medical crowdfunding in this paper.

### Social Factors in Crowdfunding Campaign

For crowdfunding research on social computing, researchers would focus on some social phenomenons or some social factors (e.g., motivations and deterrents (Gerber and Hui 2013), geographical and organizational distances (Muller et al. 2014), donor retention (Althoff and Leskovec 2015), trust amid delays (Kim et al. 2017b), and credibility factors (Kim et al. 2016)) to analyze the crowdfunding mechanism. However, the related crowdfunding studies are mainly about commercial crowdfunding, which aims at raising money for

<sup>5</sup><https://www.zgshfp.com.cn/>

	Medical crowdfunding	Poverty alleviation crowdfunding	Commercial crowdfunding
Target	Healing disease	Poverty alleviation	Accomplishing plan
Time	In a short time	In a long time	In a longer time
Platform Organizer	Companies	Governments	Companies
Reward	No reward	No reward	Reward based on donations
Fundraiser	Patient	The poor	Organization/Individual
Fee	No fees	No fees	Some fees (e.g., 5%)
Risk	No risk	No risk	High risk
Utility of information diffusion	Important	General	General
Relationship with fundraiser	Close relationship	Strange relationship	Based on interest

Table 1: The differences between medical crowdfunding, poverty alleviation crowdfunding, and commercial crowdfunding

their own commercial ventures (Tanaka and Volda 2016). For philanthropic crowdfunding, Tanaka and Volda demonstrate that the legitimacy of a philanthropic crowdfunding campaign is a central concern for fundraisers. For the medical crowdfunding, Kim et al. interview fifteen people involved in medical crowdfunding and find that support networks were larger than beneficiaries expected, with strangers offering support. Supporters offer not only monetary but also volunteering contributions, including campaign creation, promotion, and external support (Kim et al. 2017a). These works are close to our findings in this paper, but they are mainly qualitative analysis. In this paper, we devote to finding what social factors will affect the success in medical crowdfunding based on mining real log datasets. Besides, the researches above mainly study on crowdfunding in the USA (Pak and Wash 2017), while social media platforms in different countries have different credibilities and characteristics (Yang et al. 2013).

### Crowdfunding Prediction

The popularity of web content is the amount of attention that web content receives (Gao et al. 2019; Tatar et al. 2014). Users public and share the content with their friends, and several of these friends share it with their respective friends, developing a cascade of resharing can develop, potentially reaching a large number of people (Cheng et al. 2014). The particular types of popularity prediction involves tweet/microblogs (Bao et al. 2015, 2013; Cao et al. 2017), images (Zhang et al. 2018), videos (Rizoiu et al. 2017), academic papers (Shen et al. 2014) and so on. According to the timing of prediction, the popularity prediction of web content in online social networks can be divided into two types: ex-ante prediction and early prediction (Hofman, Sharma, and Watts 2017). The ex-ante prediction uses only information available before a given cascade. The early prediction will observe the progression of a cascade for some time before making a prediction (Hofman, Sharma, and Watts 2017).

Many researchers have applied machine learning to predict whether crowdfunding can be successful, but most of the research has focused on ex-ante prediction (Cheng et al. 2019; Lee, Lee, and Kim 2018). In the context of medical crowdfunding, information diffusion is an important mechanism, and widespread cases have more donation amounts or

counts. Kindler, Golosovsky, and Solomon find that some patterns at the very beginning of the campaign can be an excellent success predictor. These findings are close to some studies on popularity prediction. The studies on popularity prediction can help us better analyze, model, and predict medical crowdfunding donations. We adopt some machine learning methods on popularity prediction to our prediction of donation tasks. Our methods belong to early prediction, which usually has a better prediction performance than ex-ante prediction.

### Dataset

Our dataset is provided by Waterdrop Fundraising (Shuidi-Chou), which is one of the most popular online medical crowdfunding platforms in China. According to the reports<sup>6</sup>, from 2016 to 2019, Waterdrop Fundraising helped patients suffering from difficult and serious diseases to raise more than 20 billion yuan of medical funds in total. The number of donations has exceeded 500 million, and the monthly growth has been maintained at over 1.5 billion (CNY).

### Dataset Description

Our dataset is randomly selected from logs spanning a month in March of 2019, including a total of 28,000 cases. According to the fundraising rules, the fundraising period for a case is 30 days, and the fundraiser can end the fundraising cycle within 30 days. We removed some invalid cases whose fundraising periods are more than 30 days. In total, there are 27,618 cases. The earliest start time for these cases was 00:04:38 on March 1, 2019, and the latest end time was 23:50:05 on March 31, 2019. These cases have a total of 14M shared counts, 710K verifications, 20M donated counts, and 700M (CNY) raised money in total. With keeping data only necessary for our research, our dataset has two parts: the activity log and the detailed case information.

The log includes users' shared, verified, and donated behavior. The log of each shared activity includes shared user id (e.g.,  $A$ ), user source id (e.g.,  $B$ ), shared time (e.g.,  $t_i^s$ ), shared case id (e.g.,  $c_i$ ), and channel (e.g., WeChat). Especially, it means a user  $A$  shared a case message  $M_{c_i}$  using WeChat at time  $t_i^s$ . And the message  $M_{c_i}$  is forwarded from

<sup>6</sup><https://finance.yahoo.com/news/waterdrop-crowdfunding-topped-among-chinese-091500004.html>

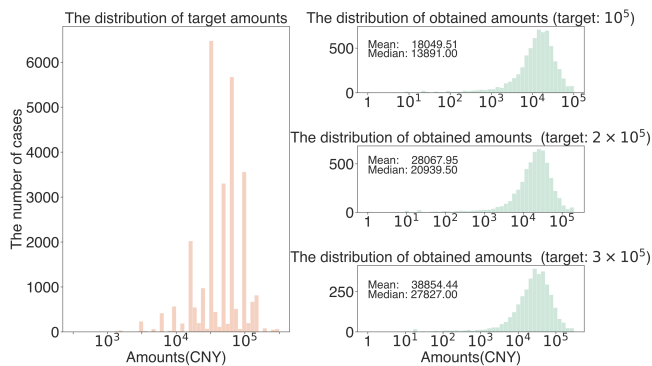


Figure 2: The distribution of fundraising target amounts and fundraising obtained amounts.

the user  $B$ . The log of each donate activity involves donated user id (e.g.,  $A$ ), donated case id (e.g.,  $c_i$ ), donated amount (e.g., 10 CNY), and donated time (e.g.,  $t_i^d$ ). It means a user  $A$  donates 10 CNY at time  $t_i^d$  for  $c_i$ . The log of each verification includes verified user id (e.g.,  $A$ ), verified time (e.g.,  $t_i^v$ ), case id (e.g.,  $c_i$ ), and relationship (e.g., relatives). It means a user  $A$  verify case  $c_i$  as the relative of the fundraiser at time  $t_i^v$ . Besides, after removing sensitive user information to protect personal privacy, we obtained detailed information for each case, including information about the fundraiser (age, gender, illness, domicile province, and hospital province), target amount, final obtained amount, and case-related text information (title and content).

### Inequality in Donations

First, we investigated whether there is an inequality in medical crowdfunding on the web (Chakraborty et al. 2019). In online fundraising, the target amount (i.e., goal) is related to the fundraiser’s expectations, reflecting the fundraiser’s needs and estimates for fundraising before spreading the fundraising information. The average fundraising goal is 172,250, and the average obtained amount is 25,732.1. The average ratio of achievement is about 0.172, which is lower than the results in GoFundMe (Berliner and Kenworthy 2017).

In Figure 2, we can find that the expectation of fundraising is mainly concentrated between 50K and 100K, but most of the fundraising cases can only raise about 10K. In the same target, different people will raise different amounts. Only very few people can achieve their fundraising goals.

### Factors Affecting Fundraising

As we mentioned earlier, the achieved ratio is relatively low, indicating that few fundraisers can finish their fundraising goals. In commercial crowdfunding, whether the goal can be achieved depends on the competitiveness of the project itself, such as commercial value and return. However, in the context of medical crowdfunding, it mainly reflects the social capital of fundraisers. For a medical fundraising campaign, there mainly exist two factors affecting the fundraising, i.e., social status factor and social network factor. On

the one hand, some people may have a higher social status, higher educations, and higher income. Some people may have more friends or their friends are more wealthy because of their social class. These factors can be viewed as social status or social profile factors. On the other hand, the social network plays a very vital role in the fundraising campaign (Tanaka and Volda 2016). Specifically, some people can make better use of their social networks to spread information more widely, getting more people’s help and raising more money. In order to understand and analyze the factors affecting the success of fundraising, we discussed the role of social status and social network separately.

### The Impact of Social Status

In this section, we focus on social status factors. The platform usually requires the fundraiser to provide some necessary information like age, name, mobile phone, diagnosis sheet, and so on. Some information will be shown in the pages on the web, and others will be assessed by the platform. Besides the mentioned necessary information, the fundraiser can edit the description information (e.g., title and content) according to their situation.

We extract the information about the fundraiser: age, gender, target fundraising amount, domicile province, hospital province, number of diseases, and text-related features (i.e., negative score, text content length, and the number of locations mentioned). Notably, we use the Chinese word segment and sentiment analysis tools<sup>7</sup> to extract the negative text score. Specifically, since the sentiment analysis tool has a limitation of the length of the analyzing text, we divide the text into chunks and then take the average sentiment score of the pieces as the final sentiment score. The negative score represents the probability of negative emotions, and it ranges from 0 to 1. If the value is less than 0.5, the text content is positive; otherwise, it is negative. Besides, we find that in the text, people usually will mention their hometown, residence, or workplace. Since that geographic information is an essential factor in fundraising, we count the number of the location mentioned in the content by matching a geographic dictionary. For extracting the number of diseases, we use some professional vocabulary to extract the disease categories<sup>8</sup>. We correlate these factors with the amount of donations using Pearson Correlation Analysis. To better show the impact of different factors on fundraising, we sort the cases according to the fundraising amount and calculate the feature statistics of the top 10% cases and the last 10% cases, marked as top10 and bottom10. From Table 2, we can find:

- In medical crowdfunding, although fundraisers are mostly male, gender is not an important factor for the amounts of donations.
- Younger fundraisers are more likely to raise more funds. This may be due to the fact that younger people are more familiar with the Internet and have more social capital.
- The higher the target amount, the more funds will be raised. This finding contradicts Wash 2013, which found

<sup>7</sup><https://intl.cloud.tencent.com/>

<sup>8</sup>[https://en.wikipedia.org/wiki/Lists\\_of\\_diseases](https://en.wikipedia.org/wiki/Lists_of_diseases)

Variable	Total	Top10%	Bottom10%	R
Gender	Male (61.3%)	Male (61.5%)	Male (61.1%)	0.002
	Female (38.7%)	Female (38.5%)	Female (38.9%)	
Age	46.96 ( $\pm 17.02$ )	35.97 ( $\pm 17.11$ )	48.60 ( $\pm 17.97$ )	-0.155**
Target amount	$10^{5.131} (\pm 0.327)$	$10^{5.401} (\pm 0.220)$	$10^{4.947} (\pm 0.426)$	0.342**
Text content length	445.70 ( $\pm 225.62$ )	572.15 ( $\pm 286.73$ )	358.36 ( $\pm 209.28$ )	0.228**
Title content length	20.18 ( $\pm 4.59$ )	20.47 ( $\pm 4.45$ )	19.44 ( $\pm 5.34$ )	0.055**
# Diseases	1.50 ( $\pm 1.14$ )	1.58 ( $\pm 1.12$ )	1.03 ( $\pm 1.05$ )	0.145**
# Locations mentioned (Province)	1.77 ( $\pm 0.89$ )	1.90 ( $\pm 0.97$ )	1.62 ( $\pm 0.93$ )	0.084**
# Locations mentioned (City)	1.79 ( $\pm 1.36$ )	2.12 ( $\pm 1.58$ )	1.47 ( $\pm 1.17$ )	0.120**
Negative score	0.32 ( $\pm 0.32$ )	0.36 ( $\pm 0.29$ )	0.33 ( $\pm 0.36$ )	0.021**

Table 2: Analysis of the factors on the fundraising information before information diffusion (\*\* $p < 0.01$ )

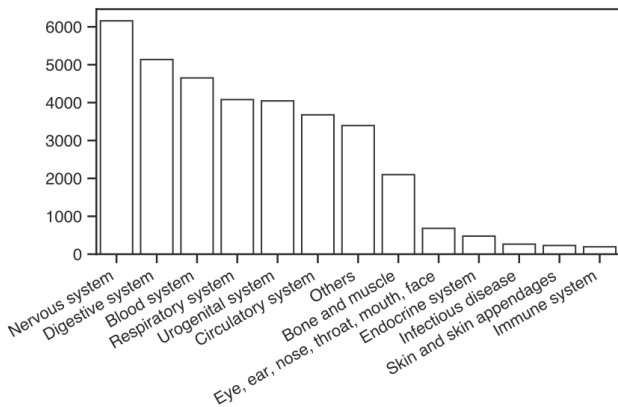


Figure 3: The barplot of disease classification.

that projects with lower fundraising goals tend to be more likely to reach their goal. The mechanisms in Waterdrop platforms may cause this. Specifically, in Waterdrop Fundraising, its fundraising rules will allow fundraisers to end fundraising without achieving their fundraising goal. There are not any refunds, even if the fundraisers may not achieve their goals. So it is usually better for fundraisers to propose higher fundraising goals.

- More detailed text information, including more geographic location information (province and city), more text description, and more extended title, are positively related to the final fundraising amounts (e.g., the correlation coefficient  $R$  of text content length is 0.228). Besides, the fundraiser who raised more money than others is concentrated in economically developed areas.
- We plot the classification of fundraising in Figure 3. By analyzing the disease, we find that most fundraising diseases are cancers, such as lung cancer, liver cancer, etc. The proportion of external trauma and infectious disease is relatively low than internal organs. This phenomenon is

consistent with the positioning of “Severe Illness” instead of “Sudden Illness”. These diseases are usually accompanied by the occurrence of other diseases (The number of diseases is large than one).

- Most of the fundraising content is positive (since the average negative score of total cases is 0.316, which is smaller than 0.5), which may be due to platform guidance. But more negative expressions will make it easier to arouse compassion and get more funds.

Generally speaking, the information submitted by the fundraiser will have a particular impact on the amount of fundraising. The fundraiser should prepare more detailed information about his situation (e.g., more location referred) to make the people know your difficulty and offer their help.

## The Impact of Social Network

In this section, we analyze the role of social network mechanisms in the process of information diffusion, including the social utility of the platform, social network verification mechanism, and close neighbors in the information diffusion cascades.

### Social Utility of the Platform

As we said before, in a case fundraising process, the platform participates in this fundraising information diffusion by showing the information on the platform. About 15% of the shared messages come from the platform website. It means that some users know and help the fundraisers by browsing the website of the platform (e.g., Love Homepage, which shows some cases on the homepage). Besides, these users also indirectly affect their neighboring users by sharing the information in their social networks.

To measure the role of the platform in the information diffusion, we build the diffusion cascade graph in action time order (e.g., if B shared a fundraising message from A, we have a link from A to B. These links are generated in time order). The starting node for fundraising is the platform (i.e., the first link is from platform to fundraiser, which

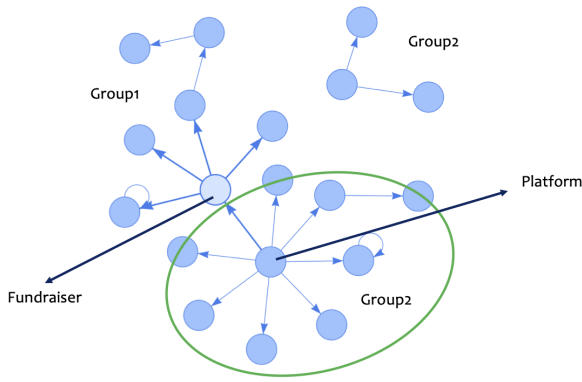
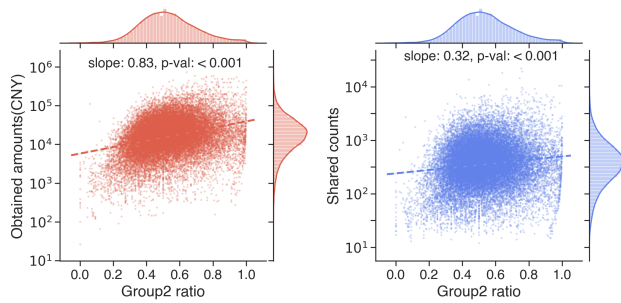


Figure 4: Illustration of our group partition. We divided the nodes which starting with the fundraiser as Group1  $G_1$  and others as Group2  $G_2$ .  $G_2$  is regarded as the social influence caused by the platform.



(a) Group2 ratio and donated amounts. (b) Group2 ratio and shared counts.

Figure 5: The relationship between  $P_{G_2}$  and obtained amounts, shared counts.

is that fundraiser share his message from the platform.). As shown in Figure 4, we mark the nodes which starting with the fundraiser or fundraiser’s successors as Group1  $G_1$ . We think this group can be social network of fundraiser can use without platforms. The nodes that cannot be started from the fundraiser are marked as Group2  $G_2$ , which is regarded as the impact of platforms. We calculate the proportion of  $G_2$  as  $P_{G_2} = \frac{|G_2|}{|G_1|+|G_2|}$ . It reflects the social utility of the platform in different cases.

We analyze the relationship between ratio of platform  $P_{G_2}$  and the number of shares and donations in Figure 5. Note that, the larger the proportion  $P_{G_2}$ , the more users come from the platform’s propagation utility. Although some user source ids are missing and  $G_1$  may be slightly different from the ground truth,  $P_{G_2}$  in all cases can still be an approximate measure of the social utility of platform in information diffusion.

From Figure 5, we can find that the proportion  $P_{G_2}$  shows a positive correlation to the number of shared counts and obtained amounts. In other words, the platform will benefit

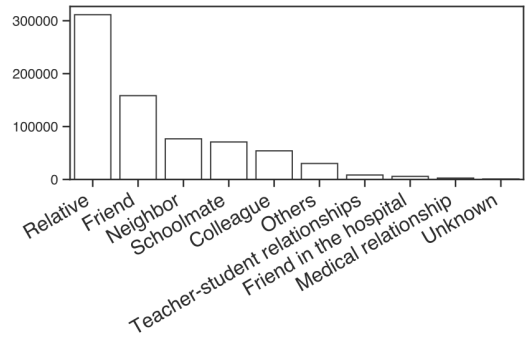
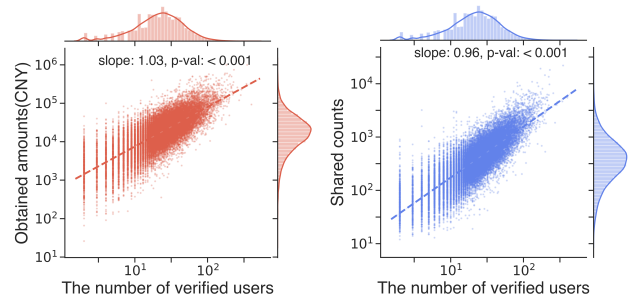


Figure 6: The barplot of the relationships between verified users and fundraisers.



(a) Verified users and obtained amounts (b) Verified users and shared counts

Figure 7: The relationship between the number of verified users and obtained amounts, shared counts.

information diffusion and donations. If someone just raises funds himself without using the platform’s diffusion utility, it will be challenging to get people’s attention and donations. The platform, as part of the fundraising on the web mechanism, has enlarged the influence and spread of fundraisers.

### Social Network Verification

During the process of the fundraiser spreads the case on his social network, he will also look for someone to confirm his fundraising case. It’s called social network verification. Those verified people usually confirm that the fundraising campaign is authentic and demonstrate the relationship with the fundraiser. We counted the relationship between these verified users and fundraisers. From Figure 6, we can see that the primary verified relationships are relatives and friends. These relationships are strong ties instead of weak ties (Granovetter 1977), indicating that people always try to find some close friends to verify their cases.

We count the verified users and correlate it with the number of shared counts and obtained amounts. Figure 7 shows that with more verified users, the amount of fundraising increased. In particular, when the number of verified users is higher than 100, the fundraising amount is usually more than 10,000, and the number of people sharing is usually higher than 1,000. The number of verified people may help the

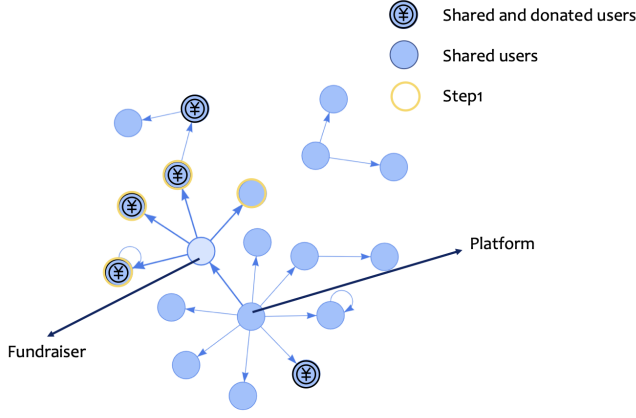


Figure 8: We define users who share the fundraising case from the fundraiser as ‘step1’. We define other users as ‘other’.

credibility of the fundraising case and further enhance the effectiveness of the information diffusion. Compared with donations and reposts, it has proven to be another important type of support for fundraisers in the medical crowdfunding (Kim et al. 2017a).

### Donation in the Information Cascade

In this section, we focus on the proportion of donations and the average amount of donations for users who have different closeness with fundraisers in the social network. We define the users who directly share the fundraising case from the fundraiser as ‘step1’ while defining other users as ‘other’. The ‘step1’ is regarded to have a more strong tie with the fundraiser than ‘other’. In WeChat, people can directly share the information of the fundraiser (i.e., shared source id is the fundraiser) only if they are friends with the fundraiser. We calculate the proportion of donation and average donation amount for ‘step1’ and ‘other’ in the information cascade. We find that the proportion of donation in ‘step1’(0.647) is significantly lower than ‘other’(0.679), which  $p\text{-Val} < 0.001$ . But the average donation(101.9) is significantly higher than ‘other’(60.5), which  $p\text{-Val} < 0.001$ . Generally speaking, the users of ‘step1’ are close to the fundraisers, but their donation proportion is lower than ‘others’. According to Chinese culture, one possible explanations is that people with close relationship are inclined to provide help to fundraisers offline, making the proportion of donations is relatively low. Meanwhile, because of the location in the social network, ‘step1’ users have a more intimate relationship than the other with the fundraiser, resulting in a large donate amount. As for those users that are far away from the fundraiser, they usually give a small amount of fundraising conservatively. This may be due to that, for the help of the friend of the friend, ‘other’ users will participate in the repost and donate, but the weak tie will lead to a small donation amount.

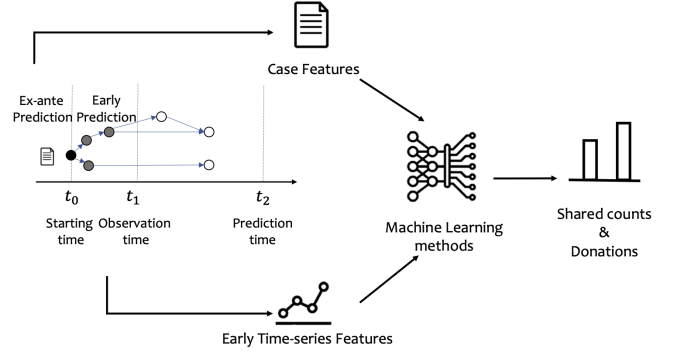


Figure 9: The black node is the original fundraiser, the gray nodes represent the shared users in the observation window, and white nodes represent unobserved ones. We select case features and early temporal features and use machine learning methods to predict the amount of sharing and donating.

### Prediction of Donations

The popularity of web content is the amount of attention that web content receives, which can be measured in many ways such as the number of views or the number of ‘likes’ (Gao et al. 2019). In medical crowdfunding, we focus on the number of shared activities and the total amount of donations like Figure 9. In social status impact, we find that the age of the fundraiser, illness, case description, and other factors are related to the total donation amounts before the information diffusion. These case features can be viewed as ex-ante features (Gao et al. 2019). In the process of information diffusion, the early temporal patterns are also related to the final shared counts and total donations. We can use some early observations to predict future popularity.

Here, we define our **prediction of donations** as following: For a medical crowdfunding campaign, given the ex-ante features  $X_c$  and the early temporal features  $X_t$  in the observation time window  $[0, T]$ , it predicts the shared counts  $y^s$  and the donated amounts  $y^d$ . For the case features  $X_c$ , we use the features mentioned before, including age, target amount, text content length, and so on. For the temporal features, we extracted the increments of the three actions (i.e., donate  $y_{\Delta t}^d$ , share  $y_{\Delta t}^s$ , and verify  $y_{\Delta t}^v$ ) in the time interval  $\Delta t$  (For donate actions, it includes donate counts  $y_{\Delta t}^{cnt}$  and donate amounts  $y_{\Delta t}^d$ ), which is

$$\begin{aligned}
 X_t &= (X_t^s, X_t^d, X_t^{d_{cnt}}, X_t^v), \\
 X_t^s &= (y_{\Delta t}^s, y_{2\Delta t}^s - y_{\Delta t}^s, \dots, y_T^s - y_{T-\Delta t}^s)^T, \\
 X_t^d &= (y_{\Delta t}^d, y_{2\Delta t}^d - y_{\Delta t}^d, \dots, y_T^d - y_{T-\Delta t}^d)^T, \\
 X_t^{d_{cnt}} &= (y_{\Delta t}^{d_{cnt}}, y_{2\Delta t}^{d_{cnt}} - y_{\Delta t}^{d_{cnt}}, \dots, y_T^{d_{cnt}} - y_{T-\Delta t}^{d_{cnt}})^T, \\
 X_t^v &= (y_{\Delta t}^v, y_{2\Delta t}^v - y_{\Delta t}^v, \dots, y_T^v - y_{T-\Delta t}^v)^T.
 \end{aligned} \tag{1}$$

In this paper, we set the time interval  $\Delta t$  to 1 hour. We sort all cascades by their publication time and take the first 80% as a training set, the middle 10% as a validation set, and the last 10% as a test set. For the length  $T$  of the observation

Prediction	Shared Counts			Donated Amounts		
Observation Time	1 day	2 day	3 day	1 day	2 day	3 day
ANN( $X_c$ )	0.9289			0.9925		
SH( $X_t$ )	0.1414	0.0694	0.0436	0.1254	0.0058	0.0338
ML( $X_t$ )	0.1247	0.0636	0.0446	0.1136	0.0533	0.0363
ANN( $X_t$ )	0.1246	<b>0.0583</b>	<b>0.0383</b>	0.1134	0.0493	0.0308
ANN( $X_t + X_c$ )	<b>0.1218</b>	0.0584	0.0388	<b>0.1110</b>	<b>0.0489</b>	<b>0.0307</b>

Table 3: Prediction of donations in Waterdrop Fundraising (mRSE, lower is better). We have bolded the highest value of each column and underlined the second value.

time window, we consider three settings, i.e.,  $T = 1$  day, 2 days, and 3 days. We choose relative square error (RSE) as our loss function, i.e.,  $f_{loss}(\hat{y}, y) = (\frac{\hat{y}}{y} - 1)^2$ , and mean RSE (mRSE) over all test data as the evaluation metric, which is used by Cao et al. 2020.

With these features, we choose typical machine learning methods to predict the donations.

- ANN( $X_c$ ): In this model, the only case features are used to train two-layer neural networks with rectified linear unit (ReLU):

$$\hat{y}^s = \text{ANN}_s(X_c), \hat{y}^d = \text{ANN}_t(X_c), \quad (2)$$

where ANN is the two-layer neural networks.

- SH( $X_t$ ) (Szabo and Huberman 2010): This model is based on the observation that the future popularity of content is linearly correlated with its early popularity. The results are predicted by:

$$\hat{y}^s = \alpha^s y_t^s + b^s, \hat{y}^d = \alpha^d y_t^d + b^d \quad (3)$$

where  $y_t^s = \sum X_t^s, y_t^d = \sum X_t^d$ ,  $\alpha, b$  are the parameters of this model.

- ML( $X_t$ ) (Pinto, Almeida, and Gonçalves 2013): It is an extension of SH model, replacing the single predictor, i.e., the cumulative popularity  $y_t$ , with multiple increments in equal-size time interval during the observation time window. The results are predicted by:

$$\hat{y}^s = \Theta^s X_t^s + b^s, \hat{y}^d = \Theta^d X_t^d + b^d \quad (4)$$

where  $\Theta, b$  are the parameters.

- ANN( $X_t$ ): We use time series features of all actions and replace the linear regression with two-layer neural networks with ReLU.

$$\hat{y}^s = \text{ANN}_s(X_t), \hat{y}^d = \text{ANN}_t(X_t) \quad (5)$$

where ANN is the two-layer neural networks.

- ANN( $X_t + X_c$ ): We add the case features into the neural networks.

These models are implemented using Pytorch (Paszke et al. 2019) with the Adam (Kingma and Ba 2014) optimizer (Learning Rate = 0.01, Weight Decay =  $1e^{-4}$ ). We train 200 epochs and select the model that performs best on the validation set for test.

Table 3 shows that only using case content will not work well for the prediction of donations and shared counts. In

contrast, the temporal features will get a good result in the task, which is consistent with previous researches (Gao et al. 2019). Longer observation windows have better results, which makes the prediction easier. ANN( $X_t + X_c$ ) achieves almost the best results in the prediction tasks. Combining both temporal features and content features are better than only using temporal features in most settings.

These experiments give us a new perspective to analyze medical crowdfunding campaigns, which uses some machine learning methods to model and predict the campaign process. Our model can be used as a new tool for the fundraisers and platform managers to monitor their crowdfunding campaigns. They can make earlier decisions that are conducive to fundraising based on the predictions of the model, to raise funds for treatment better. And these features can also better assist platform managers in designing better fundraising platforms and giving more valuable fundraising strategy suggestions.

## Conclusions and Future Work

Taken together, we empirically study the effect of different phrases in medical crowdfunding campaigns. First, we qualitatively compare the differences between different crowdfunding models, finding that information diffusion plays a vital role in medical crowdfunding when compared with traditional crowdfunding. Further, we find most people can't finish their goals. We believe that information diffusion plays an important role in medical crowdfunding, comparing traditional crowdfunding. We discuss what social factors will affect crowdfunding activity in different phases. For social status impact, we find that some personal information affects the fundraising campaign, including demographic features and text content features. For social network impact, we analyze the utility of the platform, the social network verification mechanism, and the one-hop neighborhood. Last but not least, we adopted some popularity prediction methods for predicting the shared counts and donations, which achieve a good result in the tasks.

We hope our research can shed light on medical crowdfunding, help people to raise money to save their lives, and help the organizers to improve their platform. For fundraisers, we can guide the fundraisers to describe their situation, including their geographic location information and detailed information to improve credibility. They should ask more friends to verify their cases and call for more sharing. For platforms, they can develop some monitor systems to make more early decisions to promote the campaigns. Their social network verification mechanism can be helpful for fundraising campaigns. Besides, the utility of the platform is important for medical crowdfunding. The platform should add user privacy, user fairness, and social responsibilities to the design of the platform.

In future work, we will try to study the effect of different cultures and different platforms in medical crowdfunding. Different environments and cultures will have different influences on the medical crowdfunding mechanism, and these influences objectively change the design of the platform. Comparing different medical crowdfunding in different countries is valuable for fundraising mechanism design.



## Acknowledgements

We would like to thank Waterdrop for sharing the data, Dongyang Zhang for data processing and the anonymous reviewers for their helpful comments. This work is funded by the National Natural Science Foundation of China under Grant Nos. 62041207 and 91746301. This work is supported by Beijing Academy of Artificial Intelligence (BAAI). Huawei Shen is also funded by K.C. Wong Education Foundation.

## References

- Althoff, T.; and Leskovec, J. 2015. Donor retention in online crowdfunding communities: A case study of donorschoose.org. In *Proceedings of the 24th international conference on world wide web*, 34–44.
- Bao, P.; Shen, H.-W.; Huang, J.; and Cheng, X.-Q. 2013. Popularity prediction in microblogging network: a case study on sina weibo. In *Proceedings of the 22nd International Conference on World Wide Web*, 177–178. ACM.
- Bao, P.; Shen, H.-W.; Jin, X.; and Cheng, X.-Q. 2015. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes. In *Proceedings of the 24th International Conference on World Wide Web*, 9–10. ACM.
- Belleflamme, P.; Lambert, T.; and Schwienbacher, A. 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing* 29(5): 585–609.
- Berliner, L. S.; and Kenworthy, N. J. 2017. Producing a worthy illness: Personal crowdfunding amidst financial crisis. *Social Science & Medicine* 187: 233–242.
- Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; and Cheng, X. 2017. Deephawkes: Bridging the gap between prediction and understanding of information cascades. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, 1149–1158. ACM.
- Cao, Q.; Shen, H.; Gao, J.; Wei, B.; and Cheng, X. 2020. Popularity prediction on social platforms with coupled graph neural networks. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 70–78.
- Chakraborty, A.; Mota, N.; Biega, A. J.; Gummadi, K. P.; and Heidari, H. 2019. On the impact of choice architectures on inequality in online donation platforms. In *The World Wide Web Conference*, 2623–2629.
- Cheng, C.; Tan, F.; Hou, X.; and Wei, Z. 2019. Success prediction on crowdfunding with multimodal deep learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China*, 10–16.
- Cheng, J.; Adamic, L.; Dow, P. A.; Kleinberg, J. M.; and Leskovec, J. 2014. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, 925–936. ACM.
- Gao, X.; Cao, Z.; Li, S.; Yao, B.; Chen, G.; and Tang, S. 2019. Taxonomy and Evaluation for Microblog Popularity Prediction. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 13(2): 15.
- Gerber, E. M.; and Hui, J. 2013. Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20(6): 34.
- Gierczak, M.; Bretschneider, U.; and Leimeister, J. M. 2014. Is all that Glitters Gold? Exploring The Effects of Perceived Risk on Backing Behavior in Reward-based Crowdfunding. In Myers, M. D.; and Straub, D. W., eds., *ICIS*. Association for Information Systems. ISBN 978-0-615-15788-7.
- Gierczak, M. M.; Bretschneider, U.; Haas, P.; Blohm, I.; and Leimeister, J. M. 2016. Crowdfunding: outlining the new era of fundraising. In *Crowdfunding in Europe*, 7–23. Springer.
- Granovetter, M. S. 1977. The strength of weak ties. In *Social networks*, 347–367. Elsevier.
- Haas, P.; Blohm, I.; and Leimeister, J. M. 2014. An Empirical Taxonomy of Crowdfunding Intermediaries. In *International Conference on Information Systems (ICIS)*. Auckland, New Zealand.
- Hofman, J. M.; Sharma, A.; and Watts, D. J. 2017. Prediction and explanation in social systems. *Science* 355(6324): 486–488.
- Kim, J. G.; Kong, H. K.; Karahalios, K.; Fu, W.-T.; and Hong, H. 2016. The power of collective endorsements: credibility factors in medical crowdfunding campaigns. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4538–4549.
- Kim, J. G.; Vaccaro, K.; Karahalios, K.; and Hong, H. 2017a. "Not by Money Alone" Social Support Opportunities in Medical Crowdfunding Campaigns. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1997–2009.
- Kim, Y.; Shaw, A.; Zhang, H.; and Gerber, E. 2017b. Understanding trust amid delays in crowdfunding. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1982–1996.
- Kindler, A.; Golosovsky, M.; and Solomon, S. 2019. Early prediction of the outcome of Kickstarter campaigns: is the success due to virality? *Palgrave Communications* 5(1): 1–6.
- Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lee, S.; Lee, K.; and Kim, H.-c. 2018. Content-based success prediction of crowdfunding campaigns: A deep learning approach. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 193–196.
- Lin, N.; Cook, K. S.; and Burt, R. S. 2001. *Social capital: Theory and research*. Transaction Publishers.
- Malik, M.; and Pfeffer, J. 2016. Identifying platform effects in social media data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10.
- Muller, M.; Geyer, W.; Soule, T.; and Wafer, J. 2014. Geographical and organizational distances in enterprise crowdfunding. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 778–789. ACM.

Pak, C.; and Wash, R. 2017. The rich get richer? Limited learning in charitable giving on donorschoose.org. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 8024–8035.

Pinto, H.; Almeida, J. M.; and Gonçalves, M. A. 2013. Using early view patterns to predict the popularity of youtube videos. In *Proceedings of the sixth ACM international conference on Web search and data mining*, 365–374. ACM.

Rizoiu, M.-A.; Xie, L.; Sanner, S.; Cebrian, M.; Yu, H.; and Van Hentenryck, P. 2017. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, 735–744. International World Wide Web Conferences Steering Committee.

Schwienbacher, A.; and Larralde, B. 2010. Crowdfunding of small entrepreneurial ventures. *Handbook of entrepreneurial finance*, Oxford University Press, Forthcoming .

Shen, H.; Wang, D.; Song, C.; and Barabási, A.-L. 2014. Modeling and predicting popularity dynamics via reinforced poisson processes. In *Twenty-eighth AAAI conference on artificial intelligence*.

Szabo, G.; and Huberman, B. A. 2010. Predicting the Popularity of Online Content. *Communications of the ACM* 53(8): 80–88.

Tanaka, K. G.; and Voids, A. 2016. Legitimacy work: Invisible work in philanthropic crowdfunding. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 4550–4561.

Tatar, A.; De Amorim, M. D.; Fdida, S.; and Antoniadis, P. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications* 5(1): 8.

Wash, R. 2013. The value of completing crowdfunding projects. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Yang, J.; Counts, S.; Morris, M. R.; and Hoff, A. 2013. Microblog credibility perceptions: comparing the USA and China. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 575–586.

Zhang, W.; Wang, W.; Wang, J.; and Zha, H. 2018. User-guided hierarchical attention network for multi-modal social image popularity prediction. In *Proceedings of the 2018 World Wide Web Conference*, 1277–1286. International World Wide Web Conferences Steering Committee.