User Donations in a User Generated Video System

Adele Lu Jia
China Agricultural University
Beijing, China
adele.lu.jia@gmail.com

Xiaoxue Shen
China Agricultural University
Beijing, China
xiaoxueshen@cau.edu.cn

Siqi Shen*
National University of Defense Technology
Changsha, China
shensiqi@nudt.edu.cn

Yongquan Fu
National University of Defense Technology
Changsha, China
yongquanfu@nudt.edu.cn

Liwen Peng
National University of Defense Technology
Changsha, China
pengliwen13@nudt.edu.cn

ABSTRACT

User generated video systems like YouTube and Twitch.tv have been a major internet phenomenon. They have attracted a vast user base with their many and varied contents provided by their users, and a series of social features tailored for online viewing. In hoping for building a more lively community and encouraging the content creators to share more, recently many such systems have introduced crowdsourcing mechanisms wherein creators get tangible rewards through user donations. User donation is a very special form of user relationships. It influences user engagement in the community, and has a great impact on the success of these systems. However, user donations and donation relationships remain trade secrets for most enterprises and to date are still unexplored. It is not clear at what scale are the donations or how users donate in these systems. In this work, we attempt to fill this gap. We obtain and provide a publicly available dataset on user donations in BiliBili, a popular user generated video system in China with 76.4 million average monthly active users. Based on detailed information on over 5 million videos, over 700 thousand content creators, and over 1.5 million user donations, we quantitatively reveal the characteristics of user donations, we examine their correlations with the upload behavior and content popularity of the creators, and we adopt machine-learned classifiers to accurately predict the creators who will receive donations and who will donate in the future.

CCS CONCEPTS

• **Information systems** → *Multimedia information systems*; *Collaborative and social computing systems and tools.*

KEYWORDS

User donation, User Generated Content (UGC), user generated video systems

*Shen and Jia serve as the corresponding authors of this article.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316702

ACM Reference Format:
Adele Lu Jia, Xiaoxue Shen, Siqi Shen, Yongquan Fu, and Liwen Peng. 2019. User Donations in a User Generated Video System. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3308560.3316702

1 INTRODUCTION

User generated video systems nowadays entertain over a billion users and form a billion-dollar global industry. Their success highly depends on the contents provided by their users, and a variety of social features deployed for their users to interact. To keep their proliferation, recently many such systems, for example, YouTube and Twitch.tv, have introduced crowdsourcing mechanisms and many content creators have actually attracted other users to donate [4, 5]. Being a very special form of user relationships that requires real effort to be established and maintained, user donation directly reflects and influences user engagement in the community, and therefore has a great impact on the success of user generated video systems.

However, user donation is considered trade secrets in most enterprises and to date remains an unexplored area. In this article we conduct, to the best of our knowledge, the first in-depth analysis of user donations in user generated video systems. The main motivation behind is to fill the gap between the profound and the promising role that user donation plays in real user-generated video systems and the very limited understanding of it in academia.

To this end, we have chosen BiliBili [3] as our research platform. BiliBili is a popular user generated video systems for the young generations in China. Similar to YouTube, It provides both video and social services. Since it was first launched in July 2009, BiliBili has attracted a large number of users (with 76.4 million average monthly active users for 2018) and issued their IPO for a total offer amount of 483 million dollars in March 2018 [14]. On January 15th, 2016 (much earlier than YouTube), BiliBili introduced a crowdsourcing project. Any interested creators could sign up and receive donations from other users. Unlike YouTube and Twitch.tv wherein user donation remains a trade secret and the donation statistics are not publicly available, for each creator, BiliBili displays the number of donations they have received, along with a list of the identities of the top 30 donors. By doing so, BiliBili creates a sense of friendly competition and encourages its users to donate.

Our analysis of user donations mainly consists of the following three parts:

Measuring user donations. To the best of our knowledge, we provide the first large-scale measurement and publicly available dataset on user donations in user generated video systems (Section 2). Our dataset covers 5,992,355 videos and 734,202 creators. The information we obtained includes not only basic video characteristics like the duration and the popularity, but also user activities and interactions like how users follow and donate to each other, who uploads which video and how these videos perform. For academic purposes, our dataset is publicly available upon request.

Characterizing user donations. We first quantitatively reveal the scale of user donations in BiliBili (Section 3.1). We observe over 1.5 million donations in total and over 20 thousand donations in the month of our crawling alone. While most creators in BiliBili do not receive any donations or only receive a few donations, we find that over 30% of the total donations are destined to 526 creators who have signed and declared to share content exclusively in BiliBili, indicating that exclusive user generated contents are deeply appreciated.

We then dissect the donation composition and reveal the origins of the donations (Section 3.2 and 3.3). We find that the majority (79.37%) of the donations are from viewers who have not shared any contents. Surprisingly, a considerable amount (3.99%) of donations are self-donations, which are probably used to encourage others to donate. From the donor's perspective, we find that the majority of donors only donate once whereas a few donors have returned and donated multiple times within one month.

Finally, we analyze the correlations between user donation and the upload activity and content popularity of the creators (Section 3.4). Although without qualitative analyses like surveys and interviews we cannot argue the causations, we do find that in general creators that have joined the crowdsourcing project (with and without actual donations) are more active and have shared more contents, and that for each video they share, creators with actual donations collected more views than other creators. When we compare the upload activity of the creators before and after BiliBili introduced the crowdsourcing project, we find that those who upload more frequently have received more donations.

Predicting user donations. Applying our findings, we build machine-learned classifiers to predict, without using any information on past donations, the creators *who will receive donations* and *who will donate* in the future (Section 4). On a balanced dataset our predictions achieve an accuracy of 83% and 79% for the two tasks, respectively. Our models provide insights for communities that are considering deploying crowdsourcing mechanisms and can be used to identify in advance the creators that will receive donations and the potential donors.

2 THE BILIBILI DONATION DATASET

In this section, we give a brief introduction of the ecosystem of BiliBili, and we introduce our measurement methodology and the dataset used throughout this article.

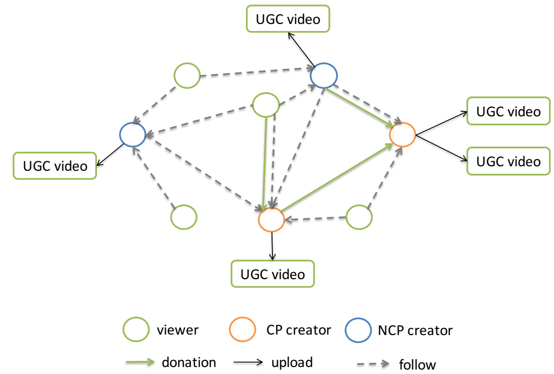


Figure 1: The BiliBili ecosystem. Three types of users include creators that have joined the crowdsourcing project (named *CP creators*), creators that did not join the crowdsourcing project (named *NCP creators*), and users who have not uploaded any contents (named *viewers*). Three types of links represent the *donation* relationships, the *follow* relationships, and the *upload* relationships, respectively.

2.1 The BiliBili ecosystem

BiliBili provides both video and social services. As in traditional user generated video systems like YouTube, users in BiliBili can share and view videos, vote and leave comments to videos, and subscribe to other users. In addition, BiliBili provides several enhanced social features, such as social network incorporation, and chat replay (named *danmu* in BiliBili) wherein the chat from the past show up right next to and on top of the video for the current viewer.

The majority of BiliBili creators are regular users with no affiliations, while a small number of the creators are *signed creators*, i.e., creators who sign up with BiliBili and declare to share contents exclusively in BiliBili, and *branding accounts*, i.e., organisations that share contents for promotion. On January 15th, 2016, BiliBili introduced a *Crowdsourcing Project* (CP) wherein any interested creators could sign up and receive donations from other users (possibly). By default, the number of donations the creators received, in total and for the current month, are highlighted in their home-pages, along with a list of the top donors.

Figure 1 shows an overview of the BiliBili ecosystem.

2.2 The BiliBili donation dataset

BiliBili identifies each of its video and each of its users with a unique number in the increasing order. Each identifier corresponds to a webpage with detailed video or user information that is publicly accessible and can be obtained with web crawlers. To give a sufficient observing period, we focus on creators that have joined BiliBili before May 2017. In total, we have obtained 136,375 and 597,827 creators that have and have not joined the crowdsourcing project, which we name *CP creators* and *NCP creators*, respectively.

For each of the CP creators, we obtain information on (i) the total number of donations he received, (ii) the number of donations and the list of the top 30 donors in the month of our crawling (April 2018), and in addition (iii) the gender, the register time, the list of users that he follows, and the creator type (regular, signed, or

Table 1: Basic statistics of the BiliBili donation dataset.

#CP creators total/signed/brand	136,375/526/325
#CP creators male/female	57,837/23,067
#NCP creators	597,827
aggregate #donations (total)	1,561,655
aggregate #donations (recent month)	51,960
#donation relationships (recent month)	25,265
#following relationships	10,543,151
#videos	5,992,355
aggregate video length	239 years
aggregate viewing time	5.37 million years
aggregate #views	50 billion
mean #views	8.468

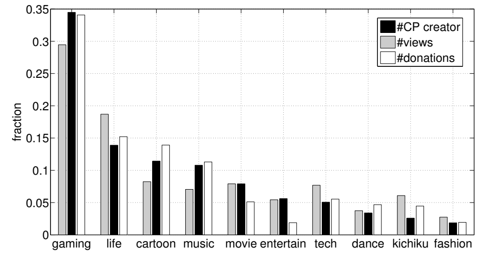


Figure 2: Fraction of CP creators, views, and donations for each category. BiliBili provides 17 categories and here we show the top 10 categories in terms of the number of CP creators. In total, it covers 97% of all the CP creators.

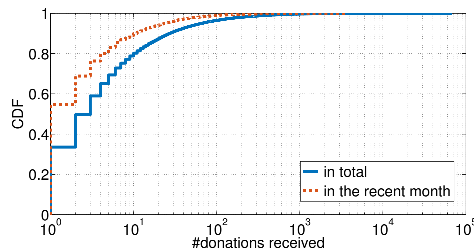


Figure 3: CDFs of the number of donations received in total and in the recent month. Note that the horizontal axis is in log scale.

brand). Overall, the CP creators have aggregated over 1.5 million donations in total and over 50 thousand donations in the month of our crawling alone. A very small fraction of them are signed creators (526) and brand accounts (325). For creators who choose to reveal their gender (which is unknown by default), we find 57,837 male creators and 23,067 female creators, respectively.

We obtain the upload activity and the content popularity of both CP and NCP creators through crawling the video pages. Our crawling was carried out in April and May 2018. To give a sufficient time for each video to collect its popularity, we have only considered videos that were uploaded at least one month ago, i.e., until March, 2018, since according to many studies on user generated video systems [6, 7, 10], video popularity rarely changes one month after the upload. In total, we have obtained detailed information on 5,992,355 videos including, for each video, (i) the uploader, (ii) the duration, (iii) the category (e.g., gaming, life, music, etc.), and (iv) the popularity attributes including the number of views, the number of favorites, the number of danmus, and the number of comments it collected.

The basic statistics of our datasets are depicted in Table 1. To give context regarding the types of the content, we briefly investigate the video category and the categories of the CP creators. Here, we choose the category of the videos that a creator has uploaded most as the category of the creator. Overall, 34.75% CP creators have uploaded videos solely from one category and the dominant category of 63.54% CP creators exceeds half of their uploads. Figure 2 shows the fraction of CP creators that fall into each category, as well as the fraction of the aggregate number of views and donations collected within each category. The gaming category, including game replays and derivatives, attracts the largest fraction of creators, views, and donations.

3 UNDERSTANDING USER DONATIONS IN BILIBILI

In this section, we reveal the basic characteristics of user donations in BiliBili.

3.1 The number of donations received

We begin by revealing the number of donations received by the 136,375 CP creators. Overall, the CP creators have aggregated 1,558,932 donations in total and 50,667 donations in the month

Table 2: Statistics on the number of donations received in total and in the recent month for different groups of creators. *User base* **shows the number of creators that have received at least one donation in each creator group.** *Fraction* **shows the fraction of total donations aggregated by each creator group.**

total	**user base**	**fraction**	Q_1	**mean**	Q_3	**max**
regular	51,164	60.91%	1	17	7	13,300
signed	517	30.17%	65	910	807	39,900
brand	262	8.78%	6	307	122	17,800
month	**user base**	**fraction**	Q_1	**mean**	Q_3	**max**
regular	5,648	57.93%	1	5	3	1,020
signed	317	34.25%	2	55	28	3,740
brand	85	6.52%	2	39	20	1,160

of our crawling alone. This result quantifies that the crowdsourcing project is running actively and is well accepted by BiliBili users.

At the individual level, however, we find that only 51,989 (38.12%) CP creators were able to actually receive any donations, among which 6,057 (11.65%) managed to receive donations in the recent month. On average, they collected 30 donations in total and 8 donations in the recent month, respectively. As further depicted in Figure 3, we observe disparities for both donation measures. Particularly, 201 creators and 66 creators have attracted over 1,000 donations in total and over 100 donations in the recent month respectively, whereas 17,457 creators have received donations only once.

Creator type. Among the 136,375 CP creators, we find 526 *signed creators*, i.e., creators who have signed up and declared to

Table 3: Donation and donator composition

donation composition	from CP creators	from NCP creators	from viewers	self-donation
count	726	3,447	20,053	1,009
fraction	2.87%	13.64%	79.37%	3.99%
donor composition	**CP creators**	**NCP creators**	**viewers**	
user base	136,375	597,827	NA	
donor count	1,669	3,312	19,293	
donor fraction	1.22%	0.28%	NA	

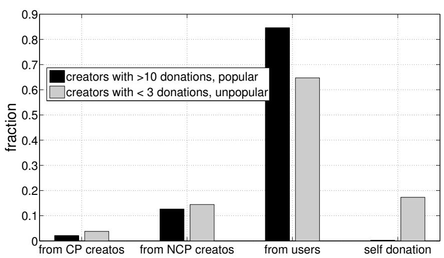


Figure 4: Donation composition: the fraction of donations received in the recent month from different groups of users for popular creators (with more than 10 donations) and unpopular creators (with fewer than 3 donations).

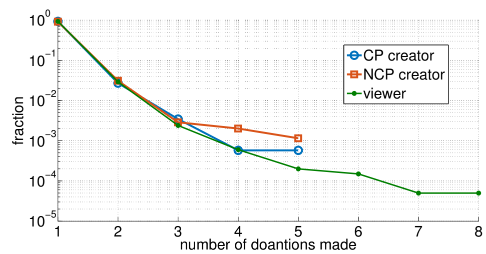


Figure 5: Donor composition: the fraction of donors by the number of donations made in the recent month. Note that the vertical axis is plotted in the log scale. In total, 96.55% donors have donated once and 3.45% donors return for a second donation within one month. Particularly, 5 donors have donated to more than five creators in the recent month and they are the viewers.

share contents exclusively in BiliBili, and 325 *brand accounts*, i.e., organizations that share contents for promotion. Intuitively, they are expected to be more professional than regular creators with no affiliations. As shown in Table 2, clearly signed creators and the brand accounts receive much more donations than regular creators. Surprisingly, while both being relatively more professional, signed creators attract roughly three times more donations than the brand accounts. And although being a minority, they have accumulated over 30% of the total donations, suggesting *deep appreciations of BiliBili viewers towards exclusive contents.*

3.2 Where do the donations come from?

The previous section has revealed that donations are conducted very actively in BiliBili. Here, we examine the donation relationships and reveal the origins of the donations. BiliBili displays for each creator the top 30 donors in the recent month. In total, we find that 6,057 creators have received donations in the month of our crawling, among which 5,816 (96.01%) have received no more than 30 donations, i.e., we have captured all the recent donation relationships for these creators and the top 30 donors (representing the most supportive ones) for the rest.

Donation composition. Depending on the origins, we decompose donations into four categories, namely donations received *from CP creators*, *from NCP creators*, *from viewers*, and *self-donations*, respectively. Reciprocal donations are rare in BiliBili. Among the 6,057 creators, we only find 8 user pairs that have donated mutually. Overall, the four donation categories each represents 2.87%, 13.76%, 79.37%, and 3.99% of the total donations, respectively. Not surprisingly, most of the donations are made by the viewers, possibly due to the large user base of viewers compared to creators. Interestingly, we find a non-neglectable fraction of donations are self-donations. We conjecture that users are encouraging others to donate by “dropping the first coin” like the street artists usually do.

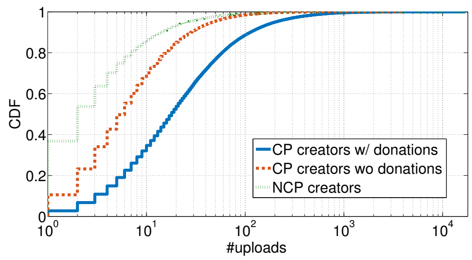
To take a closer look, we differentiate the following two groups of creators, i.e., creators with fewer than three donations (*unpopular*, in terms of donations) and creators with more than 10 donations (*popular*). Figure 4 shows the donation composition of these two types of creators, respectively. The major differences in the donations they received are that (i) unpopular creators have performed a lot more self donations; and (ii) popular creators achieve a much larger fraction of donations from viewers. It seems that “dropping the first coin” is not as effective as the unpopular creators thought.

Donor composition. The above analysis reveals the composition of donations received by each individual creator. Here, we take one step back and examine at the community level the composition of the donors. Our dataset captures in total 136,375 CP creators and 597,827 NCP creators, among which we find that 1,669 CP creators and 3,312 NCP creators have performed donations, representing 1.22% and 0.28% of the corresponding user base, respectively. Clearly, CP creators are more supportive and engaging in the crowdsourcing project.

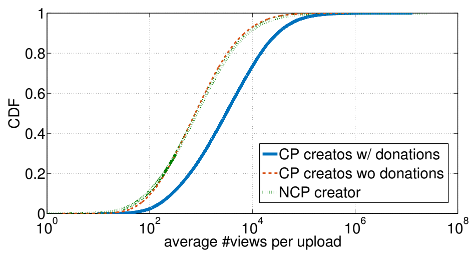
Table 3 summarizes the basic statistics for the above analysis.

3.3 Donor returns?

Finally, we focus on the donors and reveal whether donors return to make another donation. Figure 5 shows the fraction of donors that donate once to 8 times (the maximum) in the month of our crawling. Overall, 96.55% donors have donated once and only 3.45% donors return for a second donation within one month. Separating CP creators, NCP creators, and viewers, we find that a slightly higher



(a) number of uploads



(b) average number of views per uploads

Figure 6: Creator activity and popularity. Here we have divided the creators into three groups, i.e., CP creators who have received at least one donation, CP creators who did not receive any donation, and NCP creators.

fraction (3.90%) of NCP creators have returned, nevertheless, only viewers have returned for more than 5 times. The low donor return rate is possibly due to the small observing period in our experiment. For our future work, we plan to follow the creators for a longer period for verification.

3.4 Correlations with user engagement

In previous sections, we have revealed the characteristics of user donations. Here, we further analyze their correlations with user engagement. We focus on two fundamental user engagements in BiliBili, namely the creator's willingness to share new contents and the viewer's participation reflected by the content popularity (e.g., number of views). Particularly, we seek to quantitatively answer the following two questions:

1. *How are user donations correlated with the upload activity and the content popularity of the creators?*

2. *Do creators change their upload behavior after BiliBili introduced the crowdsourcing project?*

Q1: To answer the first question, we divide the creators into three groups, namely CP creators with donations, CP creators without donations, and NCP creators. We begin by examining the upload activity for creators in each group. As shown in Figure 6(a), CP creators with donations are the most active in uploading. On average they have uploaded 56 videos whereas for NCP creators it is 7.19. Interestingly, CP creators with no donations, i.e., those who have joined the crowdsourcing project but have not received any donations, are more active than NCP creators, i.e., those who did not join the crowdsourcing project.

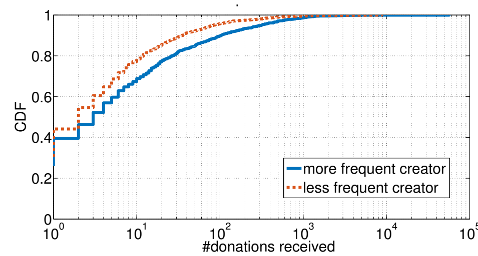


Figure 7: CDF of the number of donations received by creators that upload more and less frequently since the crowdsourcing project was introduced.

The above results quantify the correlations between joining the crowdsourcing project, receiving donations, and the upload activity of the creators. The causation, on the other hand, is difficult to argue quantitatively. It is possible that joining the project alone stimulate creators' willingness to share new contents. It is also possible that active creators are in the first place more willing to join the crowdsourcing project. Further analysis will depend on qualitative analysis such as surveys and interviews, which we leave for our future work.

Figure 6(b) further shows the CDFs of the content popularity, in terms of per-upload number of views, collected by the three creator groups. We find that CP creators with donations are in general more popular while CP creators with no donations and NCP creators achieve similar popularity. This result quantifies the correlation between viewers' participation and receiving donations, but different from the above analysis, simply joining the project does not make a difference.

Q2: To answer the second question, we compare the upload frequency of CP creators before and after BiliBili introduced the crowdsourcing project. We focus on the 4,008 CP creators who have uploaded at least 10 videos for both periods, and find that 2,785 creators have uploaded less frequently (with a longer average interval between uploads) since the crowdsourcing project was introduced, whereas 1,223 creators have uploaded more frequently. Naming them *less frequent creators* and *more frequent creators* respectively, in Figure 7 we show the distributions of the number of donations that they receive. we find that more frequent creators in general tend to get more donations.

Intuitively, there exist many possible reasons for the change in upload activities. Internally, creators could simply get bored or more engaged to the community. Externally, they could be stimulated by the feedback of the viewers (views, donations, etc.). It is possible that their high activity level exposes them to more potential donors, or that receiving donations encourage them to be more active. Similar to the **Q1**, our dataset can only quantify the correlations between them. Arguing the causation requires qualitative analysis such as surveys and interviews which we leave for our future work.

4 PREDICTING USER DONATIONS

Having gained valuable insights from the previous section, in this section we predict user donations using standard machine learning techniques. We consider two prediction tasks, namely, based on the upload and the social activity of the creators, to predict (i) *which creator will receive donations* and (ii) *which creator will donate* in the following month. Both tasks shed light on community maintenance. Particularly, the former one reveals by whom and to what extent users are attracted to the community. The latter one helps identifying potential donors whose donation will be a tangible reward for the creators.

Experimental setup. To this end, we first took a snapshot of the BiliBili community and obtained the upload activities of the creators and their content popularity until the end of March 2018. Then we focused on the 136,375 CP creators captured in our dataset and observed them for one month. During this observation period, we recorded 6,220 creators that have received donations and 4,981 creators that have donated to them. We label them as positive examples for the above-mentioned prediction tasks, respectively.

Based on previous analysis, we extract three groups of features including creator attributes (C), upload activities (U), and the follow graph properties (G). Features for the former two groups have been extensively studied in Section 3. In addition, we include features extracted from the follow graph, which is constructed by the following relationships, to test whether structural properties provide additional information on the prediction tasks. All features are summarized in Table 4.

We experimented with a variety of classification algorithms—logistic regression, SVM, and random forests—and found their performance similar. Hence all results reported here were obtained using SVM. We use balanced training and test sets containing equal numbers of positive and negative examples. For each experiment, we run 5-fold cross-validation and report the area under the receiver operating characteristic (ROC) curve (AUC).

Results. The prediction results are shown in Table 5. We have a number of interesting findings as follows.

First, consistent with the analysis in Section 3, the upload activities and content popularity of the creators (U) are very informative for predicting whether they will receive donations in the future (Task 1), however, they are not as effective for predicting who will donate (Task 2). Creator attributes, on the other hand, provide more information for Task 2. Secondly, for Task 1, the follow graph features are comparable and sometimes even more useful, showing a strong correlation between the social status of the creators and user donations. Thirdly, combining all features together achieves the best performance, and predicting who will receive donations is relatively easier than predicting who will donate: the former one achieves an AUC of 0.83 whereas for the latter task it is 0.79.

More specifically, the top three most informative features for predicting who will received donations (Task 1) are the number of followers of the creator, the number of views collected, and the number of videos uploaded previously by the creator, among which the number of views are negatively related while the rest are both positively related to Task 1. While it seems natural that creators with a large number of followers and/or have uploaded many videos are more likely to received donations in the future, it is surprising to find out that the number of views collected by the creators in fact has a negative influence on whether they will receive donations. We conjecture that viewing a video does not directly reflect user appreciation. Users may simply be exploring, or they may dislike the video and will quickly turn it off. For predicting which creators will donate (Task 2), the top three most informative features are the user type (CP or NCP creators), the number of views collected, and the number of users that creators follow. All the three features are positively related to Task 2. Different from the result for Task 1, here, the number of views creators received has a positive influence on whether they will donate to other creators. We conjecture that many creators consider receiving a large number of views as a reflection of being well accepted by the community, and they might be willing to "return the favor" in some forms including donation.

Table 4: Prediction features. We consider three groups of features and their combinations including creator attributes (C), upload activities (U), and follow graph properties (G)

group	feature description
C	gender, type, register time
U	number of uploads number of views, shares, comments, and danmus collected
G	number of followers and followees clustering coefficient, and PageRank score

Table 5: Performance evaluation (based on AUC) of predicting which creators will receive donations (Task 1) and which creators will donate (Task 2).

feature	Task 1	Task 2
C	0.5967	0.7458
U	0.8089	0.6581
G	0.7915	0.7193
$C + U$	0.8221	0.7552
$C + U + G$	0.8304	0.7852

Discussion. In the above prediction tasks, we did not use any information on the past donations, which will clearly improve the performance of our models, for the reason that we seek to infer user donations solely from the upload and the social activities of the creators. In this way, we provide insights for communities that are considering deploying crowdsourcing projects and our models can be used to identify in advance the creators that will receive donations and the potential donors. It should also be noted that for our prediction tasks we have omitted viewers, who have not shared any contents and therefore we have very limited information on them. Clearly, there exist many possible ways to improve our prediction models and also to include viewers, for example, through incorporating finer-grained network features learned from network representation learning models that are shown to perform well in node classification and link prediction tasks [13]. The key to this method is to propose new models that can cope with the heterogeneous multi-view relationships in BiliBili. We leave this as our future work.

5 RELATED WORK

We summarize related work within each research topic our work covers as follows:

User generated video systems. User generated video systems like YouTube and Twitch.tv have been extensively studied before. Cha *et al.* presented a comprehensive analysis of the popularity distribution and the time evolution of UGC video requests and their implications [6]. Ding *et al.* analyzed in-depth the behaviors of YouTube uploaders [9]. Gill *et al.* investigated YouTube from the perspective of YouTube traffic [11]. They examined YouTube usage patterns, file properties, and transfer characteristics. Kaytoue *et al.* provided preliminary characterizations on Twitch. They analyzed the dynamics of game spectators and proposed models for predicting video popularity [18]. Wattenhofer *et al.* analyzed the correlations between the popularity of YouTube videos and the properties of various social graphs created among the users [24]. In our previous works, we compared Twitch with other systems and investigated their repositories and user activities [16], and we analysed the content popularity in BiliBili and leveraged social features for user activity prediction [15, 17].

Different from the above studies, our analysis of video systems focuses on a very special form of user relationships, i.e., user donation, which to the best of our knowledge is still unexplored.

Crowdfunding systems. On the other hand, user donation in crowdfunding platforms, wherein entrepreneurs solicit funding in order to bring their business plans, have been analyzed before, ranging from predicting the success of crowdfunding campaigns [8, 12, 20, 25] to investor and project recommendation [2, 21, 22], group recommendation [23], donor retention [1] and competition modelling [19]. They mainly rely on probabilistic generative models or manual feature engineering (based on profile and social features) to build machine-learned classifiers to predict the success and the potential investors of the projects.

The above studies mainly focus on crowdfunding platforms in the context of raising money for commercial projects, wherein donors are reimbursed by receiving interests or by pre-ordering the products. Being a user generated video system, BiliBili provides a completely different context, and moreover, donors in BiliBili do not expect any real return except for some videos and a friendly community. Thus, the donor dynamics and the donation relationships are expected to be different are worth investigating.

6 CONCLUSION AND FUTURE WORK

In this work, we conducted an analysis and presented the first publicly available dataset on user donations in user generated video systems. Based on statistics on over 130 thousand creators who have joined the crowdsourcing project in Bilibili, we investigated the dynamics of user donations and we applied our findings to accurately predict future donations.

We have a number of interesting findings. First, among the 130 thousand creators, a few hundred signed creators, those who declared to share exclusively in BiliBili, have accumulated over 30% of the total donations, showing deep appreciations of BiliBili users towards exclusive contents. Secondly, the majority (around 80%) of the donations are from viewers and a considerable amount (4%) of donations are self-donations. Thirdly, we observe that joining the crowdsourcing project (even with no donations) is correlated with the upload activity but not with the content popularity of the creators. Finally, using simple features extracted from the upload activity, the popularity and the social status of the creators, we can predict with high accuracies the future donations.

As stated throughout this article, our work can be improved in a number of ways. The first would be qualitative analyses for arguing the causation on user donations. So far we have quantified (strong) correlations between user donations and the upload activity and the content popularity of the creators. It would be beneficial for the communities to understand the motivations behind, so that specific incentive policies could be designed to encourage users to donate, and creators in return will be motivated to share more. Secondly, we observe that a large fraction of donations come from viewers, on whom however we have very limited information and therefore cannot make predictions on their future donations. Relationships that they involve in, for example, social and donation relationships, would be valuable supplements and could be explored based on network representation learning models. The key to this method would be proposing suitable models that can incorporate the heterogeneous multi-view relationships in BiliBili.

ACKNOWLEDGEMENT

This work was partially supported by the National Science Foundation for Young Scholars of China (NSFYSC) No. 61502500 and No. 61602500.

REFERENCES

[1] T. Althoff and J. Leskovec. 2015. Donor Retention in Online Crowdfunding Communities: A Case Study of DonorsChoose.org. In *Proceeding of the 24th International World Wide Web Conference (WWW'15)*.
[2] J. An, D. Quercia, and J. Crowcroft. 2014. Recommending investors for crowdfunding projects. In *Proceeding of the 23th International World Wide Web Conference (WWW'14)*.
[3] BiliBili. 2017. https://www.bilibili.com.
[4] Twitch Official Blog. 2017. https://blog.twitch.tv/twitch-affiliate-program-launches-today-first-invites-going-out-9c020994a042.
[5] YouTube Official Blog. 2017. https://youtube.googleblog.com/2017/01/can-we-chat-hello-super-chat.html.
[6] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. 2007. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Internet measurement conference (IMC'07)*.
[7] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. 2009. Analyzing the Video Popularity Characteristics of Large-scale User Generated Content Systems. *IEEE/ACM Transactions on Networking* 17, 5 (2009), 1357–1370.
[8] J. Chung and K. Lee. 2015. A long-term study of a crowdfunding platform: Predicting project success and fundraising amount. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*.
[9] Y. Ding, Y. Du, Y. Hu, Z. Liu, L. Wang, K.W. Ross, and A. Ghose. 2011. Broadcast Yourself: Understanding YouTube Uploaders. In *Proceedings of the 5th Internet Measurement Conference (IMC'11)*.
[10] F. Figueiredo, F. Benevenuto, and J.M. Almeida. 2011. The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of the 4th ACM international conference on Web search and data mining (WSDM'11)*.
[11] P. Gill, M. Arlitt, Z. Li, and A. Manhanti. 2007. YouTube Traffic Characterization: A View from the Edge. In *Proceedings of the first Internet Measurement Conference (IMC'07)*.
[12] M. Greenberg, K. Hariharan, E. Gerber, and B. Pardo. 2013. Crowdfunding Support Tools: Predicting Success & Failure. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI'13)*.
[13] W.L. Hamilton, R. Ying, and J. Leskovec. 2018. Representation Learning on Graphs: Methods and Applications. In *arXiv:1709.05584*.
[14] BiliBili IPO. 2018. https://www.nasdaq.com/markets/ipos/company/bilibili-inc-1047286-86197.
[15] Adele Lu Jia, Siqi Shen, Shengling Chen, Dongsheng Li, and Alexandru Iosup. 2017. An Analysis on a YouTube-like UGC site with Enhanced Social Features. In *Proceedings of the 26th International World Wide Web Conference (WWW'17 Companion)*.
[16] Adele Lu Jia, Siqi Shen, Dick Epema, and Alexandru Iosup. 2016. When Game Becomes Life: The Creators and Spectators of Online Game Replays and Live Streaming. *ACM Trans. on Multimedia Computing, Communications, and Applications* 12, 4 (August 2016).
[17] Adele Lu Jia, Siqi Shen, Dongsheng Li, and Shengling Chen. 2018. Predicting the implicit and the explicit video popularity in a User Generated Content site with enhanced social features. *Computer Networks* 140 (2018), 112–125.
[18] M. Kaytoue, A. Silva, and C. Raissi. 2012. Watch me Playing, I am a Professional: a First Study on Video Game Live Streaming. In *Proceedings of the 12th International*

World Wide Web Conference (WWW'12 Companion).
[19] Y. Lin, P. Yin, and W.C. Lee. 2018. Modeling Dynamic Competition on Crowdfunding Markets. In *Proceeding of the 27th International World Wide Web Conference (WWW'18).*
[20] C.T. Lu, S. Xie, X. Kong, and P. S. Yu. 2014. Inferring the impacts of social media on crowdfunding. In *Proceedings of the 7th ACM international conference on Web Search and Data Mining (WSDM'14).*
[21] V. Rakesh, J. Choo, and C.K. Reddy. 2015. Project Recommendation Using Heterogeneous Traits in Crowdfunding. In *Proceeding of the 24th International World Wide Web Conference (WWW'15).*
[22] V. Rakesh, J. Choo, and C. K. Reddy. 2015. Project recommendation using heterogeneous traits in crowdfunding. In *Proceedings of the 9th International AAAI Conference on Web and Social Media (CWSM'15).*
[23] V. Rakesh, W.-C. Lee, and C. K. Reddy. 2016. Probabilistic group recommendation model for crowdfunding domains. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM'16).*
[24] M. Wattenhofer, R. Wattenhofer, and Z. Zhu. 2012. The YouTube Social Network. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'12).*
[25] A. Xu, X. Yang, H. Rao, W.-T. Fu, S.-W. Huang, and B. P. Bailey. 2014. Show me the money!: an analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI'14).*