

Who Will be The Winner in a Raffle Draw: Evidence From Weibo

Siyuan Liu, Enqi Yang, Zhipeng Liang

Sun Yat-sen University

Abstract

We consider two main parties in a Weibo raffle game, including the launcher, who launch the draw, and the participants, who retweet or like the Weibo with the purpose to win the raffle prize. Motivated by the recent controversial 'Sicong Wang's Weibo Raffle' event and the lack of research on such raffle draw activities, we are interested in the online raffle draw mechanism and conduct empirical study to tend to answer the following issues:

- Whether the Weibo raffle filtering algorithm is fair to all the participants?
- How to maximize the market influence of an online raffle draw activity ?

For the problem about fairness, we detect the raffle algorithm by conducting hypothesis tests and logistic regression to exam the relationship between prize winning probability and different properties of participants' accounts. The result also provides us theoretical support to select variables in the later section. Furthermore, time-series analysis is conducted as supplements for explanation of linear regression. Results provide Weibo accounts engaged in raffle draw activities with useful insights: participants can better evaluate the activity and launchers can efficiently design the game to reach a win-win situation.

Keywords: Weibo, raffle draw, fairness, logistic regression, online promotion

1. Introduction

Originating with the launch of Twitter back in 2006, the Microblog medium was emerging in China as early as in 2007 as a series of Twitter clones. The year 2009 saw a new evolution of the Microblog in China when the Internet giant sina.com introduced Sina Microblog, simply referred to as Weibo. On this social platform, information in the form of less than 140 words is called tweets, and information spread by users forwarding tweets is a practice known as retweeting. Weibo allows people to interact much the way they do face-to-face, honestly and authentically. Nowadays, Weibo has over 500 million registered users including verified public figures, top brands, news media, as well as a mass of grass-roots.

It is well known that Weibo raffle draw activities have become an important and efficient channel of enterprises' online promotions, especially on commemoration days or at the end of each year. These promotional methods are tend to require users to retweet the promotional messages or @ a certain number of their friends. Therefore, the users friends can then retweet and make the content viral, especially if the message creator is a celebrity. The

winners selected by the Weibo raffle draw algorithm will be provided with lucky prizes such as cashes, iPads or luxury cars freely. Through such interactive online promotion, the launchers can spread brand culture, create awareness, and more importantly promote trust and loyalty with followers, most of whom are their potential customers. (For example, the 'Finding Koi' activity launched by Alipay)

However, lack of regulations on Weibo platform has generated serious problems such as the prevailing of 'Zombie fans'. It is the term used for fake followers that lurk on Chinese social media as an infestation, hitting Weibo and the platform's credibility. They are mainly a group of internet ghostwriters paid to post online comments with particular content. Typically, these accounts only re-post news and are less active in generating original content or posting photos. Therefore, they are filtered by the raffle algorithm. Several Weibo raffle results with surprising unbalance on gender have already greatly aroused public scrutiny and controversy on the draw process and its algorithm.

To solve this problem, a typical and valuable case under fire is discussed in this paper. The story begins when Chinese e-sports club Invictus Gaming, known as iG, claimed Chinese mainlands first world championship in League of Legends (LoL). To celebrate the event, Sicong Wang, the investor, held a raffle for Weibo users who reposted his tweet. He pledged to give RMB 10,000 apiece to a total of 113 winners. The tempting raffle soon went viral not only among the countrys e-sports fans but among the more general Weibo users looking to get a piece of the action. The new has been retweeted over 22.74 million times as of November 11th. Surprisingly, the story took a different turn when 113 raffle winners were announced on November 11th, among whom only one is identified as male.

After the initial hype of the event died away, doubts about the authenticity and complaints of spamming from Netizens started to become more and more popular. According to Microblog CEO Gaofei Wang's response through his personal Weibo, the top principle for the raffle algorithm is to avoid giving the prize to accounts run by bots. Therefore, any user account would be directly filtered from the drawing process or get a lesser chance to win the raffle. The CEO tried to dispel the public suspicion on the bias of raffle mechanism against men and Android users through the logic that male users are at a disadvantage since they are usually not as active as female users in terms of content creation on Weibo.

In spite of the controversy of the raffle algorithm, it is an undeniable fact that the iG raffle hit, which was retweeted 10 billion times within hours, has become one of the most successful Microblog online promoting activities, and Wang, whose followers throve from 16 million to more than 33 million, is the greatest winner in the whole event. During roughly 4 days, more than 20 million people shared his tweet and joined in this draw. These inspired us to further explore what features of the raffle activity contribute to the success of online marketing promotion.

We conclude the introduction by giving an informal description of our models. The next section is mainly from the perspective of raffle participants from Weibo based on their properties and probability of winning the draw. It aims to detect whether the raffle algorithm behind the mechanism actually filtered zombie fans and whether the bias against men and Android users exist through hypothesis testing and logistic regression. Moreover, it simultaneously provide theoretical basis for properties selection in the later model.

Then a multiple linear regression model was conducted to evaluate the marketing engagement and features of 80 recent online raffle game. It provides deeper insights into the

impact of different-scaled activities. This permits the activity launchers better design and target the appropriate customer groups so that the goal of promoting the brand can be achieved. To explain the odd feature of activity derived from linear regression, we turn to a dynamic perspective of Weibo raffle draws using time-series analysis. And the trend tells us two effects that affect the engagement. The empirical study results can provide valuable implications for more efficient raffle draw design as well as strong regulations.

The remainder of this study is organized as follows. In Section 2, we briefly review literatures. Then, the detection of followers' properties and winning probability using hypothesis test and logistic regression is described in Section 3, followed by the multiple linear regression model of evaluating marketing effect based on the engagement of the activity and time-series analysis on Weibo raffle draw activities. Section 4 is the discussion of the results. Finally, the limitations and future research are concluded in Section 5.

2. Literature Review

As mature microblogging systems, Twitter and Weibo have been studied extensively overseas and domestically. Our study mainly derives from three streams: online raffle draw, the detection of zombie fans and prediction of retweeting.

While there have been relatively abundant studies on marketing of microblog by overseas and domestic scholars, there is little literature on the issue of online raffle draw. Chen et al. (2005) proposed a practical, anonymous, and publicly verifiable raffle protocol for use on the Internet which integrates cryptology such as public key infrastructure, hashing chain, and blind signature. Eyal Kushilevitz (2001) provided protocols for fair online lottery and online casino games which enable to remove the trust from the lottery without resorting to another trusted third party. Different from the prior literature which mainly focus on the propagation properties, profit patterns and process design of Weibo marketing, we consider Weibo raffle draw activities, an online campaign deeply related to enterprises marketing strategy and attempt to systematically explore the mechanism of this popular online promotion tool.

Another relative topic of Weibo is about the detection of zombie fans as the prevailing microblog is increasingly targeted by spammers and other malicious users to promote aliate websites and disseminate malware. Among the earliest literature to detect fake accounts, Ghosh et al. (2011) analyzed the strategies employed by contemporary spammers in Online Social Networks (OSNs) by identifying a set of spam-accounts in Twitter and monitoring their link-creation strategies. Chu et al. (2012) conducted a series of measurements to characterize the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. Our study differs significantly from the prior ones. We investigate the problem in a specific raffle draw activity to intergrade the problem with the design and fairness of flittering algorithm and mechanism of such online draw.

Another issue of great scientific significance and economic value is about the prediction of retweeting on Weibo. Many studies have been conducted to identify the influence factors of retweet behavior from different perspectives by exploring user-based and message-based features. Suh et al. (2010) gathered content and contextual features from 74 million tweets and use this data set to identify factors that are significantly associated with retweet rate. Zamanx et al. (2010) trained a probabilistic collaborative filter model for predicting the spread of information via retweet in Twitter network. They found that the identity of the

source of the tweet and retweeter were most important features for prediction. We link the prediction of the retweets with the evaluation of the raffle activity marketing efficiency. We explore the factors that contribute to an effective Weibo raffle draw based on our observation of both tweeters and retweeters.

3. Empirical Study

In this section, we first describe the data collection of Weibo users and then detail our observation of the account properties which are pivotal to the raffle draw problems by conducting hypothesis test and logistic regression. Our experiments include two parts: in the first part we investigate the factors affecting the probability of winning a draw. In order to examine whether there is discrimination against man and Android users, we attempt to search for statistical evidences to support the drawing rules revealed by Weibos CEO and provide explanations about the surprising result, which predicts the winning probability and more importantly to identity the winners patterns. They also serve as theoretical basis to help us select independent variables into our multiple linear regression model in the second part, where our perspective turns from participants to launchers. We try to explain one question: what are the key factors determining the engagement and marketing influence of a Weibo raffle draw. We investigate a few typical recent Weibo raffle draws and linear regression model is developed. Time-series analysis is also conducted to probe the features of variables.

3.1. Data Preparation

Here, we present the methodology used to crawl the Weibo and collect detailed user information. Since Weibo updates quickly and some limit on data access, the past data of the sharing accounts' list are not freely available. Thus, we cannot directly obtain the whole population of the event. To address this problem, our approach is to adapt an alternative sample collection method: we search that tweet using key words from it and collect all the users who shared that tweet in the valid time period (in fact there are still people sharing that tweet after the winning list was declared). Then we sample from that users list. To avoid distorting the original distribution, we approximate the true distribution by following the distribution of Weibo index (Figure 1), which is an official recorder for frequency of being searched of a specific word. The Weibo indexes we use here include 'Sicong Wang', 'IG' and 'ChouJiang'. Although none of them is referred directly to the draw itself, we find they highly correlated over the lottery period, which convincingly demonstrates that they can relatively approximate the searching frequency of this lottery draw activity. According to the Weibo index, we believe searching frequency is proportional to the sharing one. Following this distribution, we resample and randomly select 200 retweeters of that tweet and collect relevant data and information of the 113 winners in this raffle game. For each reached user, we record all its account properties available.

3.2. Evaluation of fairness

We first conduct data analysis to figure out what are the true factors determining the winning probability. Although 112 out of 113 winners are women, and most of the winners use iPhone instead of Android, these two factors seems unreliable as the true determines.

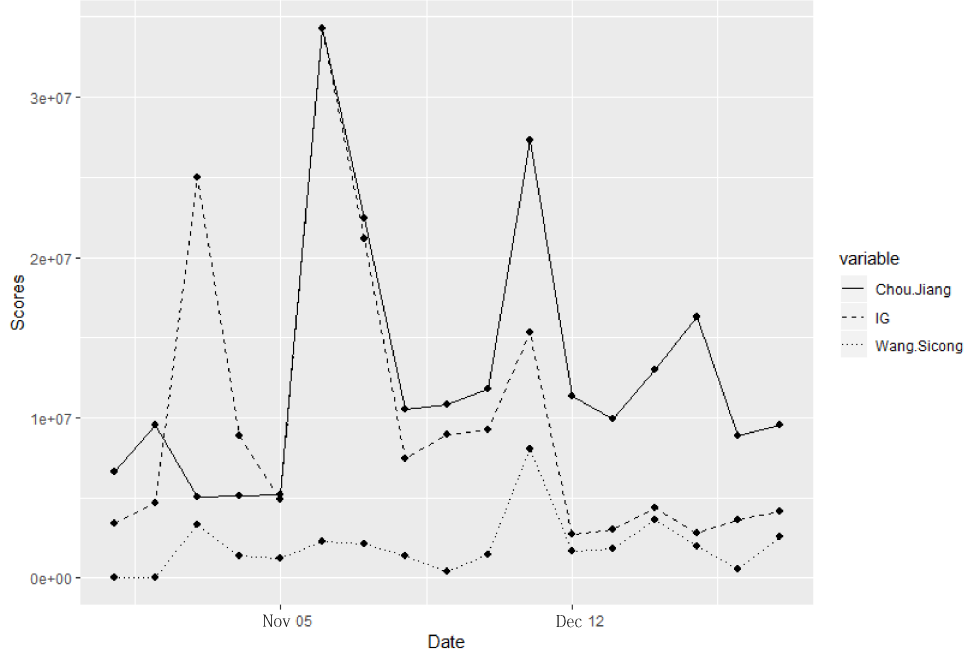


Figure 1: Weibo Index

According to Weibos CEO, lottery algorithm first removes those Zombie users before drawing a lottery. Its possible that most of Zombie users use Android and are men. The algorithm would reduce their proportion in the final winners. Therefore, we attempt to search for statistical evidences to support the drawing rules revealed and provide explanations about the surprising result.

To detect whether the raffle algorithm behind the mechanism actually filtered zombie fans and whether the bias against men and Android users exist, we first conduct hypothesis tests and then logistic regression model based on properties of the winners and other 200 common accounts from Weibo.

3.2.1. Data Description

We crawl data of 113 winners of the iG raffle draw as well as 200 ordinary participants randomly sampled from the population. For each user, we collect the following variables that reflect active levels.

Variable	Description
Level	a score that reflects users active level
Original rate	the proportion of the original Weibo issued by a user in recent one month
Have pic count	the number of the Weibo with pictures issued by a user in recent one month
Total weibo count	the number of the original Weibo issued by user in recent one month

Table 1: Variable Selection

3.2.2. Model

To better present the detecting process, we put our analysis into the following questions solved by hypothesis test.

Q1. Do the properties of winners' accounts differ from the original Weibo accounts?

In our measurement, we find out that the obvious differences actually exist, shown instinctively in the following boxplot. (We denote W as winners and A as average level)

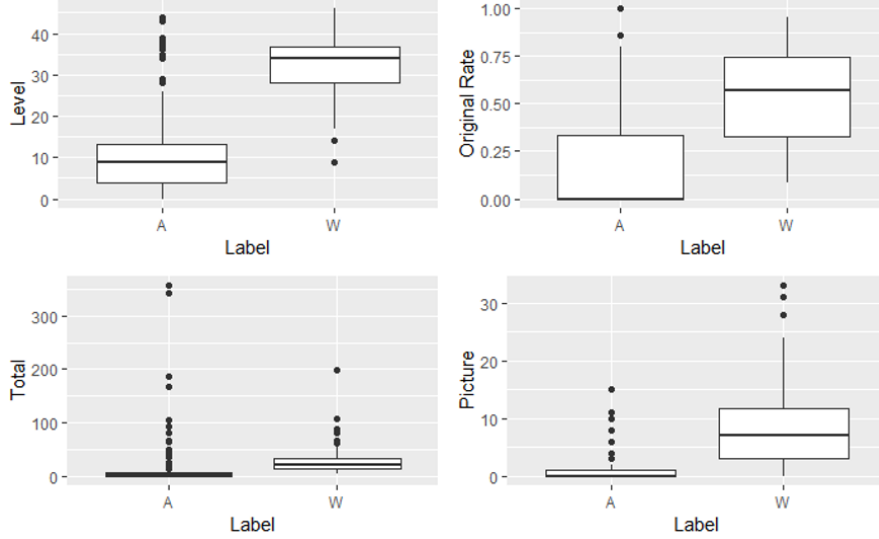


Figure 2: Boxplot of Winner-Average

	t	p-value	mean of winners	mean of average
Level***	18.377	2.2e-6	30.066038	9.265306
Total count***	7.8999	2.755e-13	27.924528	7.290816
Original rate***	10.412	2.2e-16	0.5426233	0.2065391
Weibo with pictures***	10.44	2.2e-16	7.918919	0.760000

Table 2: t-test of Winner-Average

The statistics in Table 3.2.2 may present a better insight. The first property is the account level. It indicates the active level of an account. Our measure shows that, on average the level of winners is 30, while that of common accounts is lower than 10. Thus, a high level suggests a winner and a low level implies a common account. The second property is the number of original Weibo. According to the table, about 54

Q2. Are male users as active as female users on Weibo? In our measurement, female accounts are more active than male on Weibo. As shown in ?, f defines women accounts and m defines men accounts. Intuitively, difference also exists between the male and female users' accounts.

The results from T test shows that, generally the level of female accounts is 20, far more than that of male accounts with 8. The total number of Weibo as well as Weibo with pictures posted by female accounts are significantly higher than men. On average, women post 9 total Weibo and 4 with pictures, while men only post less than one.

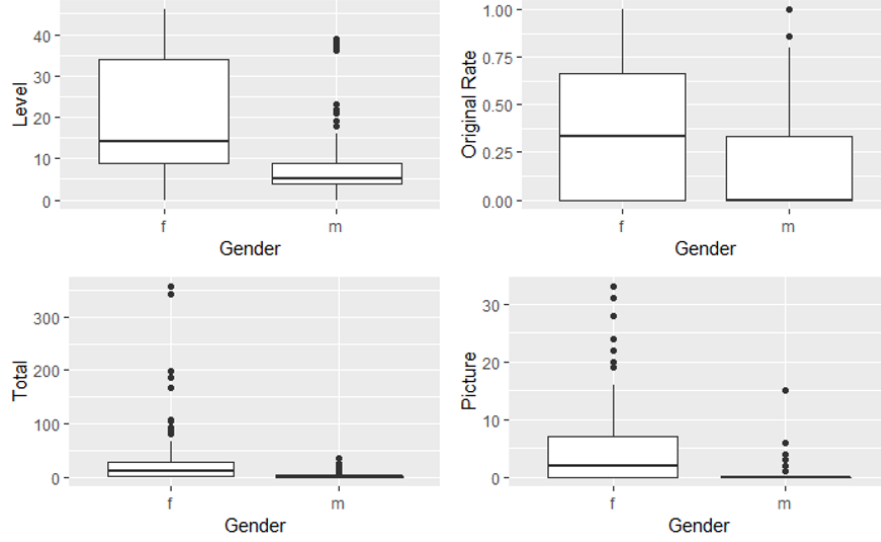


Figure 3: Boxplot of Female-Male

	t	p-value	mean of Male	mean of Female
Level***	-9.2455	2.2e-16	8.359551	20.055300
Total weibo counts***	-6.5424	1.838e-10	4.146067	23.377880
original rate***	-5.5189	5.803e-08	0.1735100	0.3842557
Weibo with pictures***	-8.3848	1.099e-15	0.5280899	4.4324324

Table 3: t-test of Female-Male

Q3. Weather iPhone users are more active than Android users? In our measurement, we have observed that Weibo accounts with iPhone users are generally more active than those with Android users. Figure 4 demonstrates the difference between users from Android and iPhone sources.

	t	p-value	mean of iPhone	mean of Android
Level***	5.4962	4.113e-8	20.42073	12.47887
Total weibo counts*	1.9074	0.02871	15.54878	11.55634
Original rate	1.4151	0.07906	0.3473376	0.2929949
Weibo with pictures***	3.6901	0.0001344	4.298780	2.014085

Table 4: t-test of iPhone users and Android users

The statistical results of hypothesis test shows that, iPhone users excess Android users in all the properties we select that reveal activeness on Weibo. For example, the average level of iPhone users is more that 20, higher than level 12 of Android users.

From the above analysis, we can conclude that different account properties really exit between winners and common users. Moreover, gender and user devices are also contributed to the surprising unbalanced result. So, what actually determines whether a user can win the raffle draw or not? To further explore the problem, we conduct logistic regression. Following is the allocation of user type with different gender and devices, providing intuitive look to

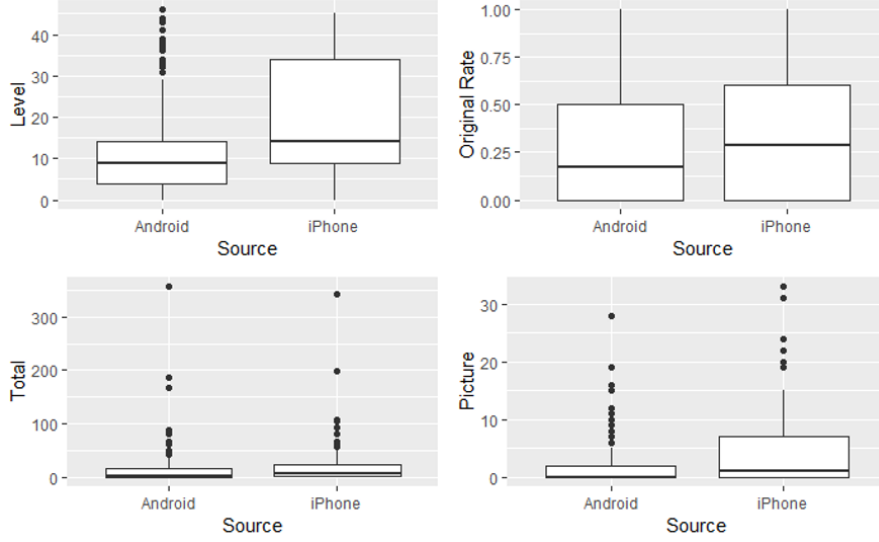


Figure 4: Boxplot of iPhone and Android

the difference.

Gender	Source	
	Android	iPhone
f	61	82
m	31	26

Table 5: Independence test between genders and mobile types

We conduct independence test between gender and mobile type and the p-value is 0.031, which enable us to conclude that different gender has different taste for phones and this contributes to the unbalance mobile types in the final results.

To summarize, the surprising result with unbalance proportions of genders and mobile types are due to the selection mechanism against the Zombie users, which prevent a large part of men from becoming candidate for their low interaction in Weibo activities. This also increase the proportion of iPhone users since women prefer iPhone than men. However, under confident level 0.05, the women-men ration in the final candidate pool must be higher than 0.84 for such a result. It seems that the selection mechanism is too hard on men, who are in fact major fans of Legend of League. Weibo should adapt more advanced algorithm to efficiently identify Zombie users to prevent mistakenly arousing people’s unsatisfactory.

3.2.3. Logistic Regression

Coefficients	Estimate	Std. Error	z value	p value
(Intercept)	-4.9870	0.60921	-8.186	2.699e-16
level	0.1442	0.02031	7.102	1.231e-12
have_pic_count	0.3444	0.07978	4.316	1.588e-05
original_rate	1.4861	0.76548	1.941	5.221e-02

From the results from logistic regression, we can conduct that level, the amount of weibo with pictures and original rate contributes statistically significantly to the winning probability. Therefore, we can conduct that the dimensions we purpose are valid.

3.3. Evaluation of marketing effect

In this section, we want to figure out how to hold a successful Weibo raffle activity. Thus, we evaluate the engagement which describes marketing effect based on the properties of launcher’s account and the activity itself.

3.3.1. Data Description

We crawl data of 80 recent Weibo raffle activities. For each activity visited, we collect its interval, number of winners, reward type and corresponding monetary value as well as the number of launcher’s followers. Variable descriptions are as follows.

Variable	Description
Interval	how long the raffle activities last
amount	how many winners of the draw
type	whether the rewards are money or gifts
money	the monetary value of prize each winner get
fans	the amount of followers of the launcher’s Weibo account
like	only require participants to like the specific Weibo
share	only require participants to share the specific Weibo

Table 6: Variable Selection

As for the dependent variables, we adapt following formula:

$$participants = \begin{cases} like\ amount, & only\ require\ to\ like \\ share\ amount, & only\ require\ to\ share \\ \min\{like\ amount, share\ amount\}, & require\ both\ the\ share\ and\ like \end{cases} \quad (1)$$

Most of the activities’ rewards are less than 1000 yuan and 5 intuitively shows that the engagement significantly differs from those with more than 1000 yuan rewards. To evaluate more accurately and efficiently, we split the sample into two kind of raffle activities: one is big prize draw with rewards more than 1000 yuan and the other is small prize draw, defined respectively as SPD and BPD in our model. Then we conduct multiple linear regression separately. We conduct stepwise regression and select three significant variables, money, amount and interval.

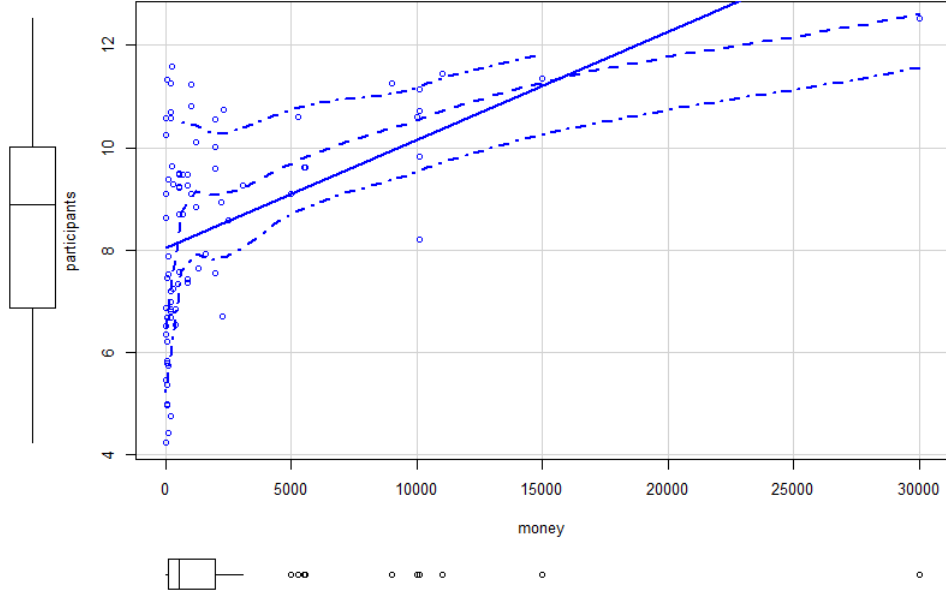


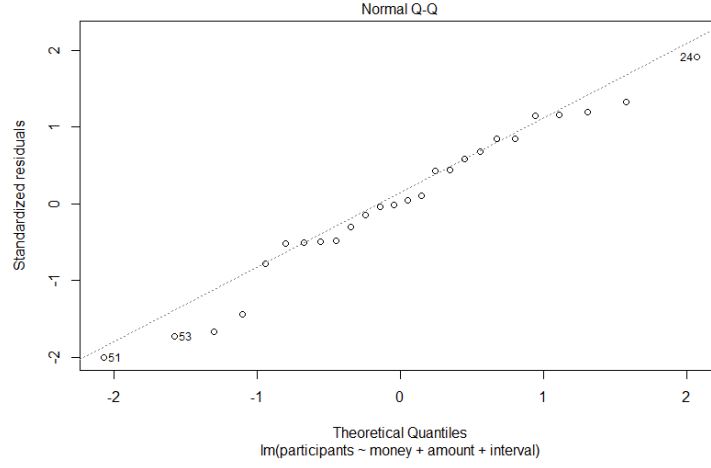
Figure 5: Scatter between participants and money

In our model, we find out that, in both BPD and SPD, money plays an important role in deciding the engagement of an Weibo raffle activity. However, BPD is generally more regular and predictable than SPD in our model. For example, in SPD the number of winners also contributes to the success of the activity, while in BPD it is less important.

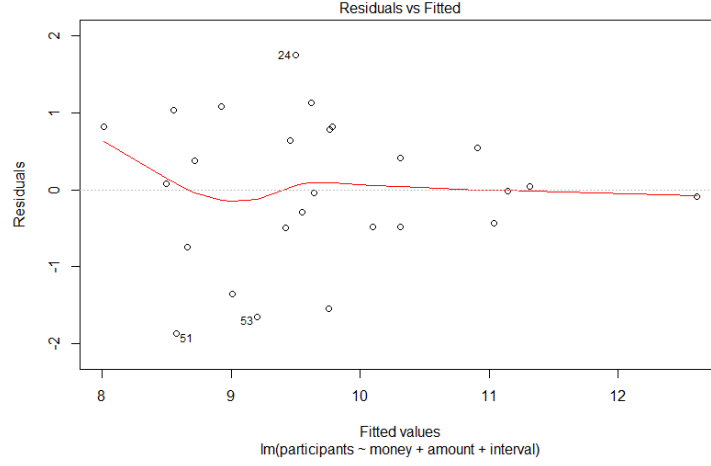
For big price draws, p-value of F test is 0.0003264 and adjusted R-squared is 0.5042, indicating the model is significant. As shown in Fig. 6(a) and 7(a), our model is good under normal QQ test and residual analysis.

	Estimate	Std. Error	t value	p-value
(Intercept)	1.032e+01	6.460e-01	15.974	1.38e-13
money***	1.483e-04	3.353e-05	4.423	0.000215
amount.	-5.746e-01	2.923e-01	-1.966	0.062085
interval*	-9.285e-02	4.230e-02	-2.195	0.039013

Table 7: Regression results of BPD



(a) QQ plot of BPD

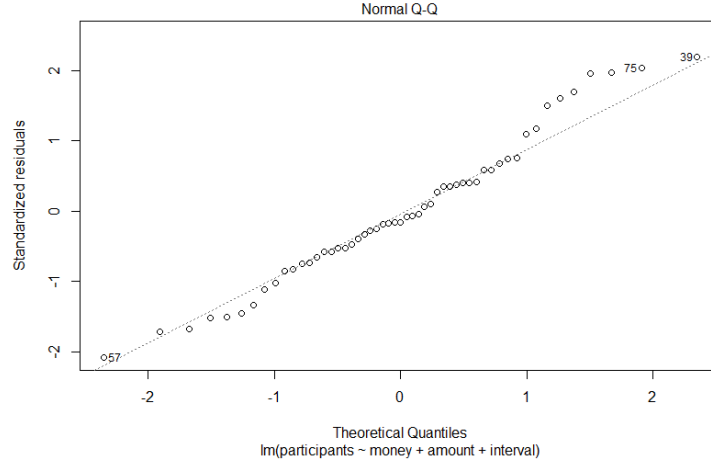


(b) Residual of BPD

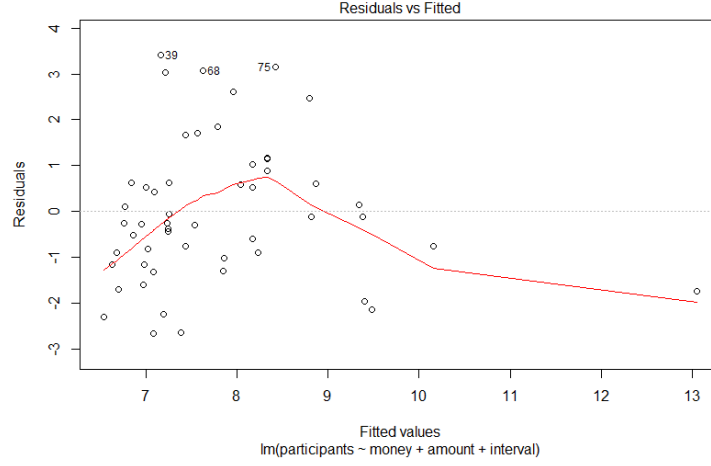
Figure 6: BPD

From the statistical result in Table 3.3.1 , both money and interval play important roles with p-value less than 0.05, while amount is not as significant as them. Moreover, we find a surprising result that the coefficient of interval is negative, indicating the longer the activity lasts, the less engagement will be in big prize draws.

For small prize draws, p-value is $8.799e-05$ and adjusted R-squared is 0.3111 of F test, which indicate the validity and significance of the model. Figure 6(a) and 6 demonstrates that our model also pass the normal QQ test and residual analysis.



(a) QQ plot of SPD



(b) Residual of SPD

Figure 7: SPD

	Estimate	Std. Error	t value	p-value
(Intercept)	6.485865	0.433219	14.971	1.2e-16
money	0.003130	0.000874	3.581	0.000784
amount	0.031658	0.007197	4.399	5.86e-05
interval	0.026879	0.051683	0.520	0.605346

Table 8: Regression results of SPD

From Table 7 we can see that money and winner amount are both very significant with p-value less than 0.01, while interval is not so significant. All the coefficients are positive, which means that the more valuable of the prize, more winners of the activity and the longer the activity lasts, the more engagement will be.

The results maybe a little strange from our daily awareness, thus we would like to provide evidence to explain why the interval becomes significance in our model and the coefficient is negative through statistical method. We imply time series analysis on 200 sample activities. The results of 200 time series plots provide us intuitive insights into the interval effect and here we select two of them as representatives presented in Figure 8. The red line describe the original engagement data and the blue one indicates the moving average. Intuitively, we can observe that, time series data of the sharing number in these two raffle activities occur at the first 20% of the period. Therefore, increasing the interval of raffle activities might not help a lot to promote the engagement and that complies with the negative coefficient of interval.

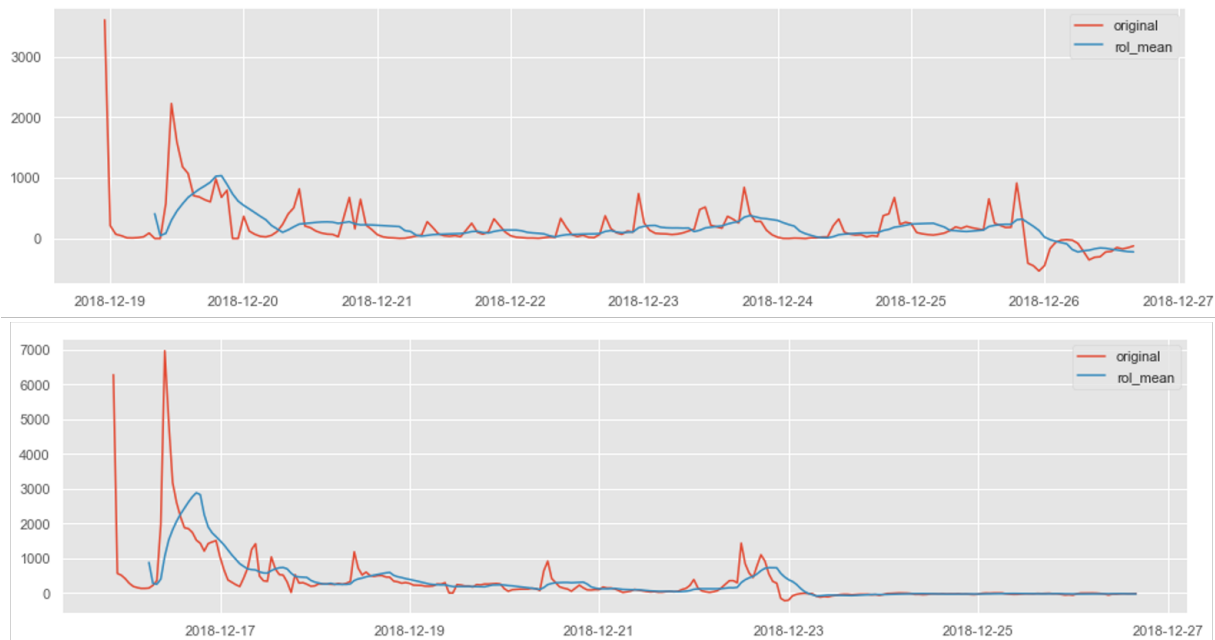


Figure 8: Timeseries data

4. Concluding Remarks

As online raffle draw becomes more and more prevalent, understanding of its mechanism and effect is essential for Weibo platform itself as well as the launchers and participants. Therefore in this study we collect data about features of Weibo users and some recent Weibo raffle draws by web crawler. We first analyze the fairness of online raffle game by conducting hypothesis test based on ever-successful Sicong Wang’s raffle game. We purpose four dimensions for determining whether a Weibo user is active or not and reveal the relationship between the patterns and the winning probability. We also develop a logistic regression model to further figure out the important factors. In the second part, we turn to launchers and investigate the decisive factors of influence of online raffle games. We identify two types of raffle activities according to their reward sizes, referring as Big Prize Draw and Small Prize Draw, and conduct multiple linear regression separately. Interval and money come as vital importance for big prize draw while money and amounts of winner contributes more to small prize draw. To provide convincing explanation of the managerial insights behind the results, we conduct time series analysis to provide evidence from dynamic prospective.

Some valuable and instructive suggestions and conclusions for different parties are listed as follows. For platform users, although publishing more Weibo could help to increase their account level, what in fact matters is to post more original Weibo, especially those with photos. The reason is that these account properties can help themselves to be recognized as active users, thus raising the winning probability. For launchers, before publishing their raffle activities, they should adopt different strategies based on monetary value of the prize. For example, for big price draw, instead of giving more winners, reducing winners and enhancing money would increase engagement and achieve better marketing influence. Moreover, publishing raffle activities just before holidays and set a rather short interval would help keep peoples enthusiasm for the reward, while for small price draw, giving rewards to more users is preferable and its better to set a relative longer time period to attract more participants. For Weibo platform, they should redesign the selection mechanism and pay more attention to male users in order not to arouse their blame and dissatisfaction, because most male users are tend to be inactive on Weibo platform.

During our research, some limitations hinder us from getting more accurate results. The first one is the limited sample size. The second is that we could provide only one explanation about the effect of interval in an less formal way. However, intrinsic property of activities and launchers can not be outlined quantitatively. Therefore, our current model fails to capture their influence and needs to be improved. More importantly, in our model there are only two parties, participants and launchers. However in reality, it is probable that a third party engages. For instance, some mediums exist and play an important role in facilitating the promotion and get paid for it. We believe that it will definitely be a interesting topic to incorporate such mediums into the analysis to investigate their roles and effect to make our model more realistic.

5. Acknowledgment

This research is supported by lecturer Lili of the course Business Data Analysis, who give us valuable suggestions and guidance to complete this work.

6. Reference

- Kushilevitz, E., & Rabin, T. (2001, April). Fair e-lotteries and e-casinos. In Cryptographers Track at the RSA Conference (pp. 100-109). Springer, Berlin, Heidelberg.
- Israel, S. (2009). *Twiterville: How businesses can thrive in the new global neighborhoods*. Penguin.
- Chen, Y. Y., Jan, J. K., & Chen, C. L. (2005). Design of a fair proxy raffle protocol on the Internet. *Computer Standards Interfaces*, 27(4), 415-422.
- Ghosh, S., Korlam, G., & Ganguly, N. (2011, March). Spammers' networks within online social networks: a case-study on Twitter. In *Proceedings of the 20th international conference companion on World wide web* (pp. 41-42). ACM.
- Ghosh, S., Korlam, G., & Ganguly, N. (2011, March). Spammers' networks within online social networks: a case-study on Twitter. In *Proceedings of the 20th international conference companion on World wide web* (pp. 41-42). ACM.
- Hutto, C. J., Yardi, S., & Gilbert, E. (2013, April). A longitudinal study of follow predictors on twitter. In *Proceedings of the sigchi conference on human factors in computing systems*(pp.821-830).ACM.
- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing* (pp. 177-184). IEEE.
- Zaman, T. R., Herbrich, R., Van Gael, J., & Stern, D. (2010, December). Predicting information spreading in twitter. In *Workshop on computational social science and the wisdom of crowds, nips* (Vol. 104, No. 45, pp. 17599-601). Citeseer.
- Artzi, Y., Pantel, P., & Gamon, M. (2012, June). Predicting responses to microblog posts. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 602-606). Associationfor Computational Linguistics.
- Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2), 351-401.