Physica A 524 (2019) 235–245

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Modeling on social popularity and achievement: A case study on table tennis[☆]

Yinghong Ma[a,*], Jiaoyang He[a], Qinglin Yu[b,c]

[a] School of Business, Shandong Normal University, Shandong, 250014, China
[b] Centre for Optimization and Data Science, Thompson Rivers University, BC, Canada
[c] School of Science, Xian Polytechnic University, Shanxi, China

HIGHLIGHTS

- The social popularity is quantified by the achievement.
- Baidu index is introduced to measure the social popularity.
- Achievements associate with the ranking, the career length, the matches' number, and the tournament's level.
- The social popularity is a timely indices relying on achievements in the match.

ARTICLE INFO

Article history:
Received 23 December 2018
Received in revised form 13 February 2019
Available online 29 April 2019

Keywords:
Baidu popularity
Achievement
Tournament value
Career length
Ranking

ABSTRACT

Reputation and fame, two measurements of social popularity, play an increasing important role in the modern society. Reputation is influenced by one's performance of success, while often divorced from her/his professional achievements. Trying to figure out the relationship between the two measures is crucial for objectively rewarding excellence in all fields, such as entertainment, sports and sciences. Baidu popularity, an index developed by the dominating searching engine company in China, is designed to reflect the reputation of a person in Chinese media and publications. In this paper, Baidu popularity is utilized to quantify the relationship of a person's social popularity and her/his occupation parameters, such as the career length, the values of tournaments, rankings, etc. Taking table tennis athletes as an example, we quantify the relationship between the success and the Baidu popularity. The data collected from web sites is used to measure the Baidu popularity of athletes. A social popularity model (SPM) is constructed by combining the logarithm of the Baidu popularity and the linear combination of the logarithms of the four observed attributes. Experiments show that the model is effective and robust, and also reveals that performance indicators are the root cause of popularity in many areas of human achievement. The real-world data confirms that the SPM model is a timely and appropriate framework to measure the reputation of table tennis athletes.

☆ This work is supported by National Natural Science Foundation of China (No. 71471106) and the Discovery Grant of the Natural Sciences and Engineering Research Council of Canada.

* Corresponding author.
E-mail addresses: yinghongma71@163.com (Y. Ma), yu@tru.ca (Q. Yu).

https://doi.org/10.1016/j.physa.2019.04.007
0378-4371/

1. Introduction

The status of fame and celebrity attract more attention in the modern society. With the rapid expansion of electronic media, fame plays a growing role in commerce, sports, entertainment, and public affairs, as well as in legal and academic spheres [1]. Different measures are used to quantify fame in different sectors. Achievement is an objective measure of a person or a group's fame in a certain domain, like the publication record of scientists or the winning record of athletes or teams [2–5]. Researchers had developed several internet-based evaluation methods for quantifying achievements [6].

The popularity of the cited scholars in the commonly used papers was quantified [7], and the number of twitter followers [8] was used to measure the visibility of scientists on social media. Google hits, the number of web pages returned in a Google search for an individual's name had been used to quantify the fame of WWI flying aces [9] and chess masters [10] as well as physicists [11].

Sports is characterized by an equally obsessive focus on popularity and achievements which strongly affect athletes' market values [12–17]. Academic research on sport is also on the rise. The relationships between the popularity of tennis' athletes and their multiple achievements indicators were investigated by accessing Wikipedia entries [18]. And the relationship between the performance and the success of football players [19] was studied quantitatively as well.

Using the search engine results or the number of followers on social media as a proxy had shown that achievements indeed drive popularity. The root of this phenomenon is that visibility and achievement are often indistinguishable [20]. However, the lacking of objective performance indicators to capture the degree of innovation or talent of a particular paper or scientist, the relationships between the achievements and the prestige are actually far from being well-understood and often controversial. Therefore, building a quantitative model to evaluate the achievements of a person has a significant value not only in scientific research but also in popular activities.

In this paper, we take table tennis athletes as an example to explore the relationships between their achievements and social popularity from the perspective of data analysis. Because table tennis is one of national sports in China, but the popular social media such as Facebook, Twitter, Wikipedia are not common in the country, we deploy Baidu index as proxy to explore the relationship between the technical performances and the social popularity of the table tennis athletes.

By studying the relationship between social popularity and performance of table tennis athletes, we examine what achievement factors affect athletes' popularity and in what ways. Taking the Baidu popularity index as a measure, the time series of achievements are evaluated by several relevant factors, such as the number of victories, professional rankings and career records. It reveals that achievements do affect athletes' popularity in positive ways.

The paper is organized in six sections: introduction shown in Section 1, data collection, variables, models, experimental results, discussions and conclusions are shown from Sections 2 to 6. Section 2 contains four parts including the data sources, the athletes, the Baidu index and Wikipedia hits. In Section 3, the four measurements/variables relevant to achievements are investigated: the tournament value, the career length, the number of matches and the ranks of the social popularity. Subsequently, a social popularity model based the observations is constructed in Section 4. The experiment results and application of the model to individual athlete are discussed in Section 5. Finally, conclusions and discussion are presented in Section 6.

2. Data

Three sets of data collected from different sources are used for the research: the first set is athletes' achievements obtained from the official website of The International Table Tennis Federation (ITTF), the second set is the athletes' attributes, and the last one is the observations of social popularity along the time recorded Baidu and athletes' Wikipedia hits.

The data choosing are briefly as: the single athletes of the top 400 were got from the ITTF official website in April 2018. Achievements data were collected from the ITTF website from 2008 to March 2018. And Baidu Index is counted once a day from January 1st, 2011.

2.1. Data sources

Achievements of the table tennis athletes are obtained from the ITTF website, where all matches played by professional male table tennis athletes from 2008 to March 2018, including names, ITTF IDs, games, opponents and collaborators in games, winnings and losings rates of the athletes, and the rankings of athletes in March and April in 2018 and more. The website of Sina Athletics Storms in China displays the Chinese name of the top players, the ranking and points in each month of the athletes from January 2001 to September 2017. The website www.allabouttabletennis.com contains the competitions and the match time of the international table tennis tournaments from the year 2011 to 2018, as well as the rankings and points of the top 100 players in each month, from September 2017 to April 2018.

2.2. The data set of athletes

We search the top 400 single players in the history of table tennis, and the top 400 players listed on the ITTF website in April 2018, it comes out a total of 564 distinct players initially. Since Baidu index is based on searching of keyword in Chinese, we use the international ranking website of Sina (a Chinese sport media) to match the Chinese name of athletes. Only 304 athletes with Chinese names were identified, and these athletes are listed on the ITTF ranking lists every month. Among them, 83 athletes are included in Baidu index. We search each athletes' Chinese name in Baidu website to confirm whether the identity and information of each athlete is correct. Excluding the athletes with different translations and partial information, there are only 37 athletes with full information. The basic information of those 37 male table tennis athletes is shown in table 4 of Appendix and we will refer it as the experimental sample I.

The record of male table tennis athletes began in the year 2008, while the Baidu index starts its coverage in January 2011. Therefore, the time interval of our data set is from January 2011 to March 2018.

2.3. Time interval Δt and time-series baidu index $B(t)$

Baidu index is calculated daily starting on January 1st, 2011. While, the lengths of tournaments are different for various competitions. The game time of a table tennis tournament usually lasts about 7 days, however, Olympic Games generally lasts 11 days. Other shorter table tennis games generally last from 3 to 5 days. In any case, the game time does not exceed two weeks. The official ranking of table tennis athletes is updated once a month. Considering the effect of delay in reporting game scores, we choose one month as our time interval (Δt) which coincides with the frequencies of updating the official ranking at the ITTF and the game lengths.

The monthly Baidu index $B(t)$ is used to measure the social popularity of table tennis athletes. We calculate the Baidu index of 37 male table tennis athletes from January 2011 to March 2018 for 87 months, and obtain 3219 data points. During the period of January 2011 to March 2018, some athletes are retired, so the number of the athletes still in games is 37 persons with 2477 data in finally.

2.4. Wikipedia page visits $W(t)$

We also collected the page visits to the athletes' Wikipedia pages as the comparison with the Baidu index. Wikipedia provides daily search volumes to the athletes' Wikipedia pages and analyze the users' browsing activities since July 2015. The data of Wikipedia pageviews of the table tennis athletes is searched for the period of July 2015 to March 2018.

Taking English name of an athlete as the keyword, among the 37 athletes identified in Section 2.3, only 34 athletes have daily pageviews. Deleting the athletes whose Wikipedia pages are created after the year 2018, and the ones retired during the period, we end up with 25 athletes and 742 data points meeting the requirements. The remaining 25 athletes are referred as samples II. The basic information of these 25 athletes is shown in Table 3 of Appendix.

3. The key variables

To evaluate achievements of each athlete, we consider several relevant factors in reflecting the different aspects of achievements: the social popularity by Baidu index $B(t)$, Wikipedia page-views $W(t)$, the tournament value $V(t)$, the career length $Y(t)$, the number of matches $N(t)$, and the professional rankings $R(t)$.

3.1. The social attentions

Social attentions are represented by the Baidu index, which is a free data service based on the search in the social media and the news, and reflects the "user's attention" and "media attention". Baidu index provides daily search volume driven by keywords in China since January 2011. We also account the page visits of the athletes' Wikipedia and compare it with the athlete's Baidu index to validate the Baidu index as a measure of social attention.

In this paper, the Chinese names of the table tennis athletes are used as keywords. We collect the Baidu index data of the athletes from Baidu.com between January 2011 and March 2018. As an illustration, we take the famous male player Jike Zhang as an example. The relationship of Baidu index and the professional ranking is shown in Fig. 1.

Labels in Fig. 1 indicate the trends of Jike Zhang's rankings and Baidu index. Labels 2 and 7 correspond to the London Olympics in 2012 and Rio Olympics in 2016, respectively. Because of the greater influence of Olympic Games around the world, people paid much more attention to the athletes. Jike Zhang won the men's single final of London 2012 with 4 to 1 over Hao Wang that aroused great social popularity which led to the peak of label 2. At the 2016 Rio Olympic Games, because of his coach Guoliang Liu's comments "Jike Zhang wake up, this is the Olympic", Jike Zhang attracted countless fans latterly by tenacious fighting spirit and professional performance. His Baidu index reach an unprecedented height, shown as label 7. It shows that the social popularity of the athletes is not only related to the official ranking, but also closely related to special incidents, the public attentions as well the achievements of the athletes.

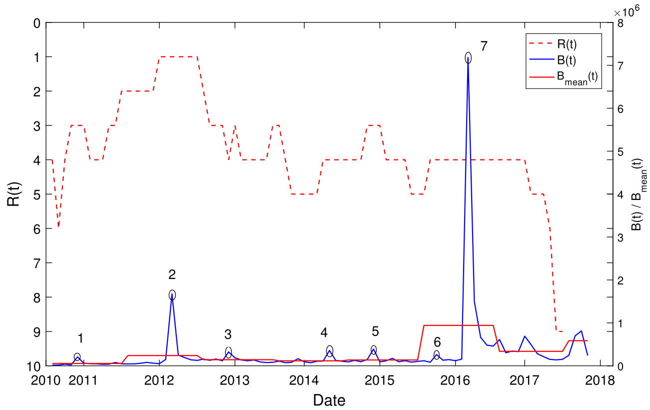


Fig. 1. The illustration of Jike Zhang's rankings and Baidu indices, where rankings $R(t)$, Baidu index $B(t)$ and the annual average Baidu indices $B_{mean}(t)$ are shown in dotted red, blue and red, respectively. The x-coordinate is the time, the left y-coordinate is rankings $R(t)$, and the right y-coordinate is the ratio of Baidu indices of $B(t)$ and $B_{mean}(t)$. The data is from January 2011 to December 2017 and the time interval Δt is one month.

3.2. The tournament value $V(t)$

The well-recognized tournaments offer points to the winners. The more points a tournament offers to the winners, the more prestige it becomes. So it is reasonable to take the points of tournament of the winner as a measurement of the achievements for athletes. The value of a tournament is the points of the winner earned. For example, the champion is given 3000 points by ITTF in the Olympics or the World Championships. So the value of the Olympics or the World Championships is set to be 3000 points. In the World Cup, the champion receives 2550 points, so the value of the World Cup is 2550 points. For example, the early peaks in Fig. 1 correspond to Jike Zhang's participation in the World Table Tennis Championships, and he got the champion. Therefore, the higher value an athlete's competition gets, the more social attention he received. The values of different types of games are shown in Table 5 of the Appendix.

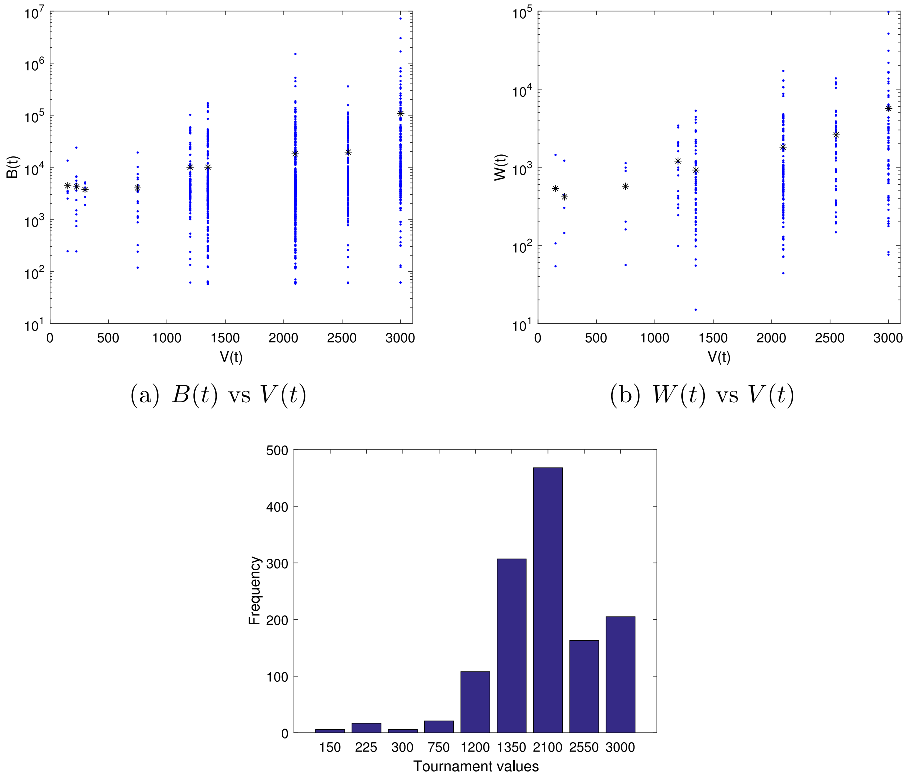
We investigate the relationships between social popularity and tournament values of athletes, shown in Fig. 2.

Figs. 2(a) and 2(b) demonstrate the trends of the tournament values and the Baidu indices or the Wikipedia page-views of athletes, respectively. The two panels show that the higher the tournament value, the higher the Baidu index the athletes receive, so does Wikipedia page-views. In Fig. 2(a), the average Baidu index drops when $V(t) = 2500$, which is probably due to the lower number of participation in tournaments with the increasing level of competition. The Fig. 2(c) shows that frequencies of athletes taking part in matches with different values are diversified. When $V(t)$ is between 150 to 2100, the higher value of the tournament is, the more athletes participate. However, when $V(t) = 2550$ or $V(t) = 3000$, the frequencies of participation are decreased rapidly since the matches in these tournaments become the most fierce competitions in the world, such as the Olympics, the World Championships and the World Cup. It also reveals that the relationships between tournament values and social popularity are much similar whether measured by Baidu index or Wikipedia page-views. Since the Baidu index possess more sample data, we choose it as the primary measure for social popularity.

3.3. The career length $Y(t)$

The career length $Y(t)$ of an athlete is defined as the duration from the first year of the official ITTF ranking to the year of athletes' retirements. That is, the career length $Y(t)$ is the number of the years that his name appeared on the official list of ITTF.

The relationships between social popularity and career lengths of athletes are shown in Fig. 3. Fig. 3(a) shows that, in the first 14 years, athletes' Baidu indices have a close relationship with how long the players have been active in this sport. The active competition time positively correlate with their social popularity. And the athletes' Baidu indices decrease as time going, when their career lengths are more than 14 years. In Fig. 3(b), the average values of the athlete Wikipedia page-views also change as career length increasing. As the most of the sample points locate in the red rectangle of Fig. 3(b), the average trend in this part is relatively more convincing. Clearly, the peak values of the average occurs at 13 and 14 years, which is basically consistent with the result of Fig. 3(a). Fig. 3(c) is the frequency histogram of career lengths. When the career length $Y(t) \leq 10$, the frequencies increase as the career lengths increase, but they decrease when $Y(t) > 10$. This can be explained that most of the career lengths of athletes are about 10 years. Figs. 3(a) and 3(b) also reveal that the table tennis athletes usually reach their career peaks at about the 14th year. So the general trends of social popularity are similar in using Baidu indices or Wikipedia page-views as proxies.



(a) $B(t)$ vs $V(t)$ (b) $W(t)$ vs $V(t)$

(c) Distribution of $V(t)$

Fig. 2. The tournament values relate to social popularity from January 2011 to March 2018. (a) The plot of Baidu index $B(t)$ versus the tournament values $V(t)$ for 37 athletes. (b) The plot of Wikipedia page-views $W(t)$ versus the tournament values $V(t)$ for 25 athletes. (c) The frequency histogram of tournament values for 37 athletes. In panels (a) and (b), the black stars indicate the averages of the Baidu index and the page-views for Wikipedia, respectively.

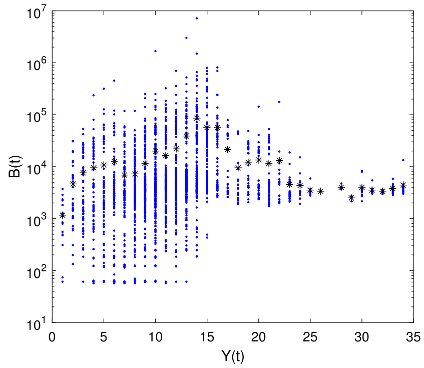
3.4. The match number $N(t)$

An athlete may participate in more than one tournament in a month, and also in different types of competitions in a tournament, such as singles, doubles, or team events. The number of matches of athletes in a tournament might reflect their achievements within the competition, and would increase their social popularity. Denote the maximum number of matches by $MN(t)$ and the total number of matches by $TN(t)$ for athletes participating in tournaments in a month. We investigate the two parameters in relations with the social popularity. The four panels in Fig. 4 demonstrate the relations of the number of matches and social popularity.

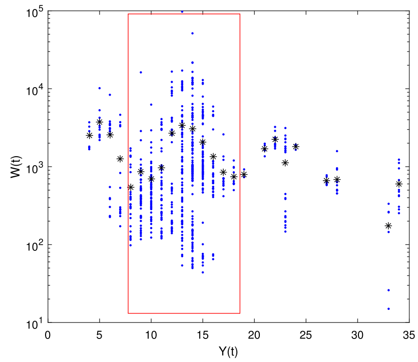
$MN(t)$ versus Baidu indices $B(t)$ or Wikipedia page-views $W(t)$ are displayed in Figs. 4(a) and 4(b), respectively; $TN(t)$ versus Baidu indices $B(t)$ is shown in Fig. 4(c). Roughly speaking, the averages of Baidu indices of athletes increase with the growth of the maximum number of matches per month. The number of athletes' Wikipedia page-views increases with the growth of their maximum number of matches per month too. $W(t)$ in Fig. 4(b) acts similarly as that of $B(t)$ in Fig. 4(a), while the two social popularity are not obviously related to the total number of matches, shown in Fig. 4(c). Therefore, we choose the maximum number of matches $MN(t)$ as a variable in the process of model building. We call $MN(t)$ the match number and denoted it by $N(t)$ in the following analysis.

Clearly, a few athletes are eliminated after one match in a competition, and most of them play two or more matches, but only a few top athletes move to the semi-finals or finals, which is demonstrated by the distribution of $N(t)$ in Fig. 4(d). The distribution is nearly normal, and a relatively small number of athletes have higher match times. As in any sport, the semi-finalists and finalists attract more attention and thus generate the most popularity. In this case, the Baidu index rises along with $N(t)$.

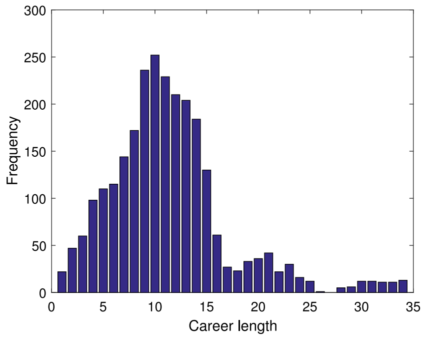
Examining the behaviors of $N(t)$ versus $B(t)$ or $W(t)$ in Fig. 4, we find that the Baidu index is more suitable as a proxy of social popularity than Wikipedia.



(a) $B(t)$ vs $Y(t)$



(b) $W(t)$ vs $Y(t)$



(c) Distribution of $Y(t)$

Fig. 3. The career lengths relate to social popularity from January 2011 to March 2018. (a) The plot of Baidu indices $B(t)$ versus the career lengths $Y(t)$ for 37 athletes. (b) The plot of Wikipedia page-views $W(t)$ versus the career lengths $Y(t)$ for 25 athletes. (c) The frequency histogram of career length for 37 athletes. In panels (a) and (b), the black stars indicate the averages of the Baidu index and the page-views for Wikipedia, respectively.

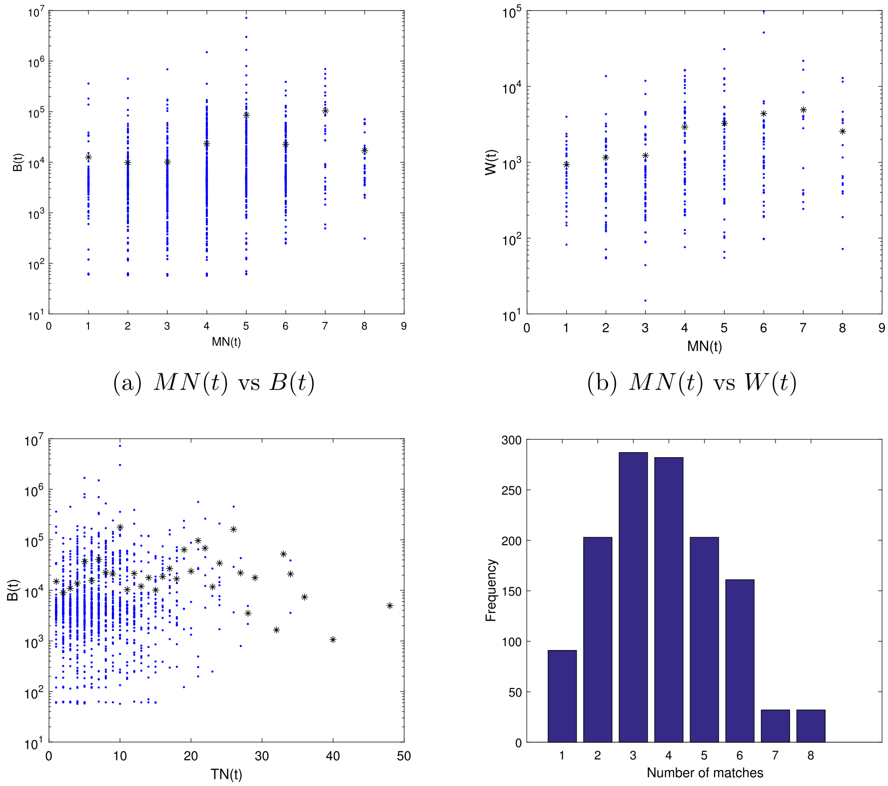
3.5. The ranking value $R(t)$

Rankings are the direct measures of athletes' achievements in a sport. We use the official rankings of ITTF, denoted it as $R(t)$. In ITTF ranking system, the best athlete is ranked as seed number 1, i.e., $R = 1$, and the second best athlete is ranked as seed number 2, i.e., $R = 2$, and so on. The lower value of $R(t)$, the better of the overall performance. The monthly Baidu indices of 37 athletes are shown in Fig. 5(a) and the monthly Wikipedia page-views of the 25 athletes is shown in Fig. 5(b).

Fig. 5(a) shows that $B(t)$ decreases as the ranking decreases when the ranking is less than 80. There are several special points affect the trend. For example, the label 1 corresponds to Guoliang Liu's comments to Jike Zhang in the Olympic Game on August 2016 which results his Baidu index extremely high for the month. The influence continues in the following month, the label 2 is still very high. We find out the reasons for the labels 3 and 4, which are due to the announcements of relationship between Jike Zhang and Tian Jing who is a well-known actress in China, and consequently increases the social popularity in 2018. The other extremely points are also due to some special events which amplify the social popularity (i.e., the Baidu index). When the ranking is more than 80, due to the small number of sample points, the Baidu indices become disorder with the increase of the ranking values in Fig. 5(a).

Fig. 5(b) is the scatter plot of the athletes' Wikipedia page-views and the rankings. It shows similar trend as that of the Baidu indices. When the values of rankings are relatively small, the Wikipedia page-views are crowded and display increasing trends as the rankings grow. While the Wikipedia page-views disperse when the rankings are relatively large. The reason might be the size of available samples. In summary, Fig. 5 shows that the social popularity of athletes increases with the decreasing of rankings.

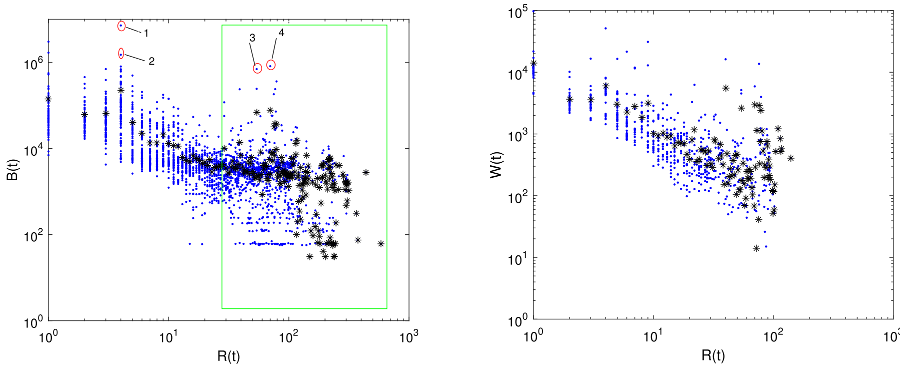
Comparing the effeteness of Baidu indices and the Wikipedia page-views as proxies of social popularity, shown in Figs. 2 to 5, both measures demonstrate the trends consistent with the changes of the athletes' professional achievements. However, the Baidu index has more sample data and behaves more consistently. So, it is more appropriate to use the Baidu index to measure the social popularity of athletes in this study.



(a) $MN(t)$ vs $B(t)$ (b) $MN(t)$ vs $W(t)$

(c) $TN(t)$ vs $B(t)$ (d) Distribution of $N(t)$

Fig. 4. The number of matches and social popularity from January 2011 to March 2018. (a) The plot of Baidu index $B(t)$ versus the maximum number of matches $MN(t)$ for 37 athletes. (b) The plot of Wikipedia page-views $W(t)$ versus the maximum number of matches $MN(t)$ for 25 athletes. (c) The plot of Baidu index $B(t)$ versus the total number of matches $TN(t)$ for 37 athletes. (d) The frequency histogram of the maximum of matches $MN(t)$ for 37 athletes. The black stars represent the averages of the social popularity in panels (a), (b) and (c), respectively.



(a) $R(t)$ vs $B(t)$ (b) $R(t)$ vs $W(t)$

Fig. 5. The rankings $R(t)$ and social popularity from January 2011 to March 2018. (a) The plot of Baidu index $B(t)$ versus the rankings $R(t)$ for 37 athletes. The ten numbered points correspond to special events with high Baidu indices. (b) The plot of Wikipedia page-views $W(t)$ versus the rankings $R(t)$ for 25 athletes. The black stars indicate the average of the social popularity in panels (a) and (b).

4. The social popularity model

The analysis in Section 3 show that the professional achievements of athletes do influence social popularity. The rankings, the career lengths, the number of matches and the tournament values are all playing roles in it. And they all display nonlinear relations with the achievements of athletes.

Table 1
Pearson correlations of the four variables.

	$1/R(t)$	$Y(t)$	$N(t)$	$V(t)$
$1/R(t)$	1.0000	0.0503	0.2326	0.1198
$Y(t)$	0.0503	1.0000	−0.0301	0.0700
$N(t)$	0.2326	−0.0301	1.0000	0.0683
$V(t)$	0.1198	0.0700	0.0683	1.0000

Table 2
p-values and β-coefficients for SPM.

	$1/R(t)$	$Y(t)$	$N(t)$	$V(t)$
p-value	0.000	0.022	0.000	0.000
β-coefficients	0.659	0.049	0.108	0.088

In order to discover the intrinsic connections, the Pearson correlations of the four variables are considered at first. Table 1 shows values of the Pearson correlations of them by computing the data set of 37 athletes.

The largest value of the correlation among the variables is 0.2326 which is between $N(t)$ and $1/R(t)$. That implied that the linear correlation between any two variables is weak. So, we seek to build a model of social popularity in all four variables.

Deploying the tool of Stochastic Impacts by Regression on Population, Affluence and Technology [21,22], which evaluates the aforementioned factors that describe their individual potential impact on global warming, we assess how the four variables mentioned in Section 3 can impact the social popularity. Here, we build a model that represents Baidu index as a nonlinear mathematical expression, named Social Popularity Model (SPM), and evaluate its suitability and robustness in empirical analysis.

$$(SPM): \; B_E(t) = a(\frac{1}{R(t)})^{a_1} Y(t)^{a_2} N(t)^{a_3} V(t)^{a_4} \varepsilon(t), \tag{1}$$

where a is a constant, a_i ($i = 1, 2, 3, 4$) are the exponents of variables $1/R(t)$, $Y(t)$, $N(t)$ and $V(t)$, respectively, which need to be determined, and $\varepsilon(t)$ is a term for white noise.

We test the fitness of the SPM model with the data sets by the ordinary least squares fitting process [23] and the standardized β-coefficients. The R-squared of the SPM model is 0.5223 and the p-values of t-test for all variables shown in Table 2: since all the values are less than 0.1, the test is accepted.

On the other hand, the four variables are measured in different terms, such as orders, years, times and points. So we evaluate which independent variable has a greater effect on the dependent variable. The standardized β-coefficients is used to measure how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable [24]. Therefore, by calculating the standardized β-coefficients for each term in Eq. (1), we evaluate how strongly each variable influences $B(t)$ by their standardized β-coefficients. Table 2 shows that the ranking is the strongest influence factor for social popularity.

We also establish four nonlinear models using three out of four variables, see Equations (5)–(8) in appendix. But the SPM model is the strongest with the R-squared value 0.5223 and Akaike Information Criterion (AIC) [25] value 3577.81 among the five models (see Table 6 in Appendix). That is, the SPM model offers the best predictive accuracy among the models we tested.

Eq. (1) can also be expressed as the following logarithmic form.

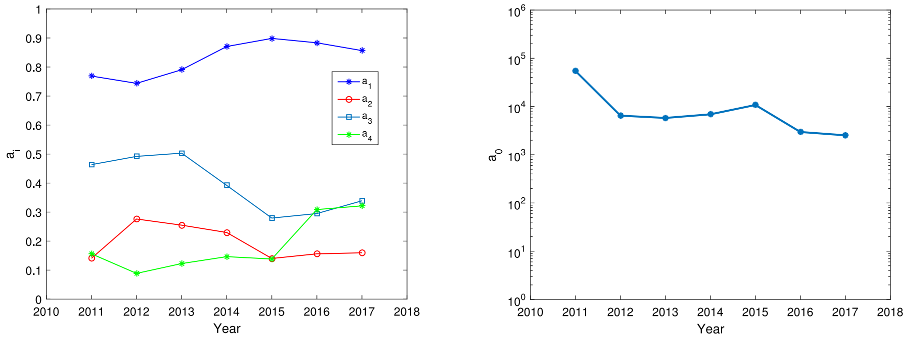
$$\log B_E(t) = a_0 - a_1 \log R(t) + a_2 \log Y(t) + a_3 \log N(t) + a_4 \log V(t), \tag{2}$$

where $a_0 = \log a + \log \varepsilon(t)$ is the logarithms of the constant term and the error term.

In fact, we will use Eq. (2) as our basic model in the rest of the paper, to conduct experiments, regression analysis and fitting assessment.

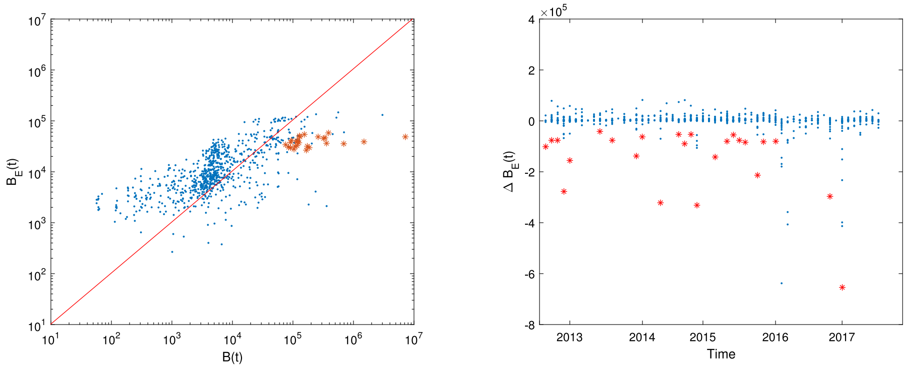
5. Experiments

To assess the accuracy of the SPM model, we compare it with other models and validate its predictive power. Taking the first two years of observations as training data, the values of a_0 is larger at first while the data sets of the first two years do not have a significant effect on it. The other four coefficients have not much fluctuation. Hence, we use the observations of the first two years, 2011 and 2012 to fit the model (2), and obtain the fitting values of a_1, a_2, a_3, a_4 and a_0, respectively. The trends of a_i are shown in Figs. 6(a) and 6(b). Approximately, $a_1 = 0.744$, $a_2 = 0.276$, $a_3 = 0.492$, $a_4 = 0.089$, and $a_0 = 6460.47$.



(a) a_i (b) a_0

Fig. 6. The coefficients of the SPM model. (a) The changes of a_i (i = 1,2,3,4) with the increase of data. (b) $a_0 = \log a + \log \varepsilon(t)$ of the model (2).



(a) $B_E(t)$ and $B(t)$ (b) $\Delta B_E(t) = B_E(t) - B(t)$

Fig. 7. Prediction of social popularity. (a) The correlations of the predication from the SPM model $B_E(t)$ and the observed data $B(t)$. (b) The values of $\Delta B_E(t)$ over time. The red stars are the data of Jike Zhang corresponding to the two panels.

Fig. 7(a) shows the predications from the SPM model and the real data $B(t)$. Notably, the most data are around the line $B_E(t) = B(t)$, which implies that the predication coincide with the observations. Fig. 7(b) shows the distribution of $\Delta B_E(t)$ over time.

In order to compare the superiority of the SPM model, we construct other two models (see Appendix, the models (3) and (4). The model (3) is a linear combination of the four variables, and the model (4) is the special case of the SPM model when all the exponents equal 1. Figures 9(a) and 9(c) in Appendix show the predications of the models (3) and (4), and Figures 9(b) and 9(d) are the differences of $B_S(t)$, $B_M(t)$ and the observations data $B(t)$, respectively. Comparing Figs. 7(b) and 9(b), 9(d), it is clear that the more points in Fig. 7(b) are clustered around $\Delta B_E(t) = 0$, which indicates that the prediction of the SPM model is more accurate.

Comparisons of the models of $B_E(t)$, $B_M(t)$, $B_S(t)$ with $B(t)$ reveal that the SPM model (1) is the best fitting function for Baidu index.

Finally, we apply Jike Zhang's data from 2011 to 2017 to the SPM model to check the predictability of the SPM. The scatter plots of the forecasting data $B_E(t)$ and the observed data $B(t)$ are in Fig. 8. The predicted data of social popularity, shown in dot red line, is smaller than the real data. The trends of the two curves are consistent, expect the spike in 2016. This large deviation is caused by the wide media coverage on a single event explained in Section 3.1. Overall, the SPM model $B_E(t)$ predicates the social popularity of table tennis athletes quite well.

The SPM model $B_E(t)$ takes the rankings, the career lengths, the number of matches, the values of tournaments together into consideration to predict the social popularity of table tennis athletes. The social popularity not only relies on their rankings, but also their career length and the levels of tournaments.

At 2009, Baidu Search becomes the most popular searching engine in China. And Google pulled out of Chinese market at that time. Therefore, we chose the Baidu popularity data from the year 2011. Because the table tennis is Chinese national sport and more popular in Asian countries, it is no surprise that the majority of TV audiences is from Asia especially from China. So choosing Baidu index from China as a proxy for social popularity of table tennis is not only because we have more data from Baidu index, but also audience base and the nationality of the top players.

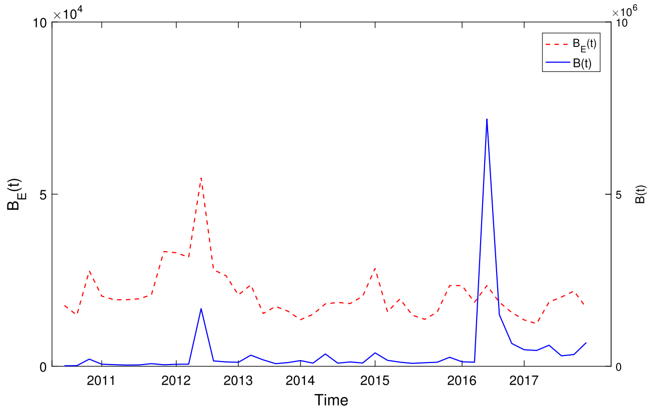


Fig. 8. Comparison between Jike Zhang's Baidu index $B(t)$ (solid line) and his performance-based predicted social popularity $B_E(t)$ (dotted line). The x-axis is times in years, the left y-axis is the forecasting data $B_E(t)$, and the right y-axis is the observed data $B(t)$.

6. Discussion and conclusions

At first sight, the social popularity seems to be a simple concept representing some form of recognition from single or continued achievements. However, many studies suggested that the social popularity is not limited to one's talent or achievements. In this study, we attempt to establish a relationship between achievements and social popularity in quantitative manner.

Due to the particularity of the mass audiences of table tennis, the data of the popularity are observable, available and acceptable to the public. By comparing Baidu indices and Wikipedia page-views, we find that Baidu index is more suitable to measure the social popularity of athletes than Wikipedia page-views.

The social popularity model (SPM) is constructed based on the four variables: the rankings, the career years, the tournament values and the matching number, to predict Baidu index of athletes. We find, in table tennis, that the players' popularity and instant social attentions can be linked to their on-court performance, in particular the spectacular events by the elite athletes in top competitions. So the extraordinary performances provide an unusual amount of visibility, which is difficult to adjust through ordinary games. However, for most outliers, some events can be ignored in our model, such as the achievements of doubles or junior tournaments, and the news impact generated being public figures outside of their professional careers.

The SPM model is possible to evaluate the other areas where performances and observations can be measured independently. For examples, since the ranking mechanisms in sports like badminton, chess or golf are similar to table tennis, we can use the same approaches to explore how professional athletes' achievements are related to their social popularity.

In other fields, such as scientific research, the prestige of scholars is closely related to their publication records and discoveries. The challenge, however, is how to separate prestige from social popularity, and how to discover impartial indicators to measure achievements.

Some further works are worth to do in the future. For example, due to the technical limitations, the data in this paper is not large enough, which requires Baidu website to gather more data; many people watches sports by mobile phones instead of TVs or PCs via broadcast from BBS and Sina. Because Baidu index is generated by the searchings on Baidu.com, to include other media exposures or venues will capture a more accurate picture of social popularity. Moreover, that the factors influencing social popularity are much more complex, so to include more factors could potentially improve the accuracy of the model. It is true there is some limitation using Baidu index to model the popularity. It is luckily that table tennis is the national sport in China and Baidu is easier to use than Wikipedia, the social popular model measures the players' achievements of table tennis.

Acknowledgments

We would like to thank the anonymous reviewers for the constructive comments and suggestions, which undoubtedly improved the presentation of this paper. We also show our appreciation for the National Science Foundation of China (No. 71471106) that supports this research.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2019.04.007.

References

[1] P. Drake, A. Miah, The cultural politics of celebrity, Cult. Polit. 6 (1) (2010) 49–64.
[2] F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: toward an objective measure of scientific impact, Proc. Natl. Acad. Sci. 105 (45) (2008) 17268–17272.
[3] F. Radicchi, Who is the best player ever? a complex network analysis of the history of professional tennis, PLoS One 6 (2) (2011) e17249.
[4] F. Radicchi, Universality, limits and predictability of gold-medal performances at the olympic games, PLoS One 7 (7) (2012) e40335.
[5] S. Mukherjee, Quantifying individual performance in cricket a network analysis of batsmen and bowlers, Physica A 393 (1) (2014) 624–637.
[6] E.D. Ramirez, S.J. Hagen, The quantitative measure and statistical distribution of fame, PLoS One 13 (7) (2018) e0200196.
[7] Y. Ding, B. Cronin, Popular and/or prestigious? measures of scholarly esteem, Inf. Process. Manage. 47 (1) (2011) 80–96.
[8] A.Z. Yu, S. Ronen, K. Hu, T. Lu, C.A. Hidalgo, Pantheon 1.0, a manually verified dataset of globally famous biographies, Sci. Data 3 (2016) 150075.
[9] M.V. Simkin, V.P. Roychowdhury, A mathematical theory of fame, J. Stat. Phys. 151 (1–2) (2013) 319–328.
[10] M.V. Simkin, V.P. Roychowdhury, Chess players' fame versus their merit, Appl. Econ. Lett. 22 (18) (2015) 1499–1504.
[11] J.P. Bagrow, H.D. Rozenfeld, E.M. Bollt, D. ben Avraham, How famous is a scientist? – famous to those who know us, Europhys. Lett. 67 (4) (2004) 511–516.
[12] S. Merritt, A. Clauset, Scoring dynamics across professional team sports: tempo, balance and predictability, EPJ Data Sci. 3 (1) (2014) 1–21.
[13] A. Clauset, M. Kogan, S. Redner, Safe leads and lead changes in competitive team sports, Phys. Rev. E 91 (6) (2015) 062815.
[14] A.M. Petersen, W.-S. Jung, J.-S. Yang, H.E. Stanley, Quantitative and empirical demonstration of the matthew effect in a study of career longevity, Proc. Natl. Acad. Sci. 108 (1) (2011) 18–23.
[15] S. Motegi, N. Masuda, A network-based dynamical ranking system for competitive sports, Sci. Rep. 2 (12) (2012) 904.
[16] P.O.S. Vaz de Melo, V.A.F. Almeida, A.A.F. Loureiro, C. Faloutsos, Forecasting in the nba and other team sports:network effects in action, ACM Trans. Knowl. Discov. Data 6 (3) (2012) 1–27.
[17] S. Herm, H.-M. Callsen-Bracker, H. Kreis, When the crowd evaluates soccer players market values: Accuracy and evaluation attributes of an online community, Sport Manage. Rev. 17 (4) (2014) 484–492.
[18] B. Yucesoy, A.L. Barabasi, Untangling performance from success, EPJ Data Sci. 5 (1) (2016) 1–10.
[19] L. Pappalardo, P. Cintia, Quantifying the relation between performance and success in soccer, Adv. Complex Syst. (2017) 1–29, arXiv:1705.00885v3.
[20] Y.-B. Zhou, L. Lu, M. Li, Quantifying the influence of scientists and their publications: distinguish prestige from popularity, New J. Phys. 14 (3) (2012) 33033–33049, (17).
[21] E.A. Rosa, T. Dietz, Climate change and society - speculation, construction and scientific investigation, Int. Sociol. 13 (1998) 421–455.
[22] R. York, E.A. Rosa, T. Dietz, Stirpat, ipat and impact: analytic tools for unpacking the driving forces of environmental impacts, Ecol. Econom. 46 (3) (2003) 351–365.
[23] J. Cohen, P. Cohen, S.G. West, L.S. Aiken, Applied Multiple Regression/correlation Analysis for the Behavioral Sciences, Lawarance Erlbaum Associates, 2003.
[24] L.D. Schroeder, D.L. Sjoquist, P.E. Stephan, Understanding Regression Analysis: An Introductory Guide, Sage Publications, 1986, pp. 31–32.
[25] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, second ed., Springer, 2002.