TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 01/11 pp313–324
DOI: 10.26599/TST.2019.9010011
Volume 25, Number 3, June 2020

Modelling the Dropout Patterns of MOOC Learners

Zheng Xie*

Abstract: We conduct a survival analysis for the viewing durations of massive open online courses. The hazard function of the empirical duration data presents as a bathtub curve with the Lindy effect in its tail. To understand the evolutionary mechanisms underlying these features, we categorize learners into two classes based on their different distributions of viewing durations, namely lognormal distribution and power law with exponential cutoff. Two random differential equations are provided to describe the growth patterns of viewing durations for the two classes respectively. The expected duration change rate of the learners featured by lognormal distribution is supposed to be dependent on their past duration, and that of the remainder of learners is supposed to be inversely proportional to time. Solutions to the equations predict the features of viewing duration distributions, and those of the hazard function. The equations also reveal the features of memory and memorylessness for the respective viewing behaviors of the two classes.

Key words: data science applications in education; distance education and online learning; evaluation methodologies

1 Introduction

Massive Open Online Courses (MOOCs) arose from the integration of education and information technologies, characterized by unlimited participation and open access via the Internet[1,2]. The effects of these courses include breaking the spatiotemporal boundary of traditional education, and contributing to balancing education resources[3,4]. There are several differences between MOOCs and traditional courses, including conditions for admission, learning motivation, teaching methods, the management of learners, and the interactions between teachers and learners[5,6]. Analyzing learning behaviors has become a hot topic in the MOOC community. There are a number of aspects to this analysis, such as the achievements of learners[7,8], the interactions among learners[9], the visual analysis of course data[10], and the assessment of courses[11].

One feature of MOOCs is a high dropout rate, which is thought to result from the diverse expectations and motivations among people taking these courses[12–15]. MOOC learners are not just motivated to pass exams or to obtain certificates. They may be only interested in understanding particular concepts or certain contents of a course[16–19], and are then likely to drop after obtaining the knowledge they were seeking. Analyzing the dropouts of MOOCs contributes to quantifying the completion and continuance of learning[20,21]. For example, how many learners will continue to learn after they have progressed to a certain point? At what rate will leaners drop out in the future? Moreover, understanding the evolutionary mechanisms underlying the dropout behavior and modelling these mathematically contributes to profiling types of learners[22] and to quantifying the effects of teaching methods and other explanatory variables on learning behaviors[23–25].

The log data of learning behaviors collected by MOOC platforms, of which viewing time is the most prominent[26], can be used to analyze the dropout rate. We adopt survival analysis for learners' viewing

• Zheng Xie is with the College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410000, China. E-mail: xiezheng81@nudt.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2019-03-30; accepted: 2019-04-04

duration on a course, with the empirical data provided by the iCourse platform (http://www.icourse163.org). The survival function of the durations describes the fraction of the learners who view a course past a given time, and the corresponding hazard function describes the dropout rate of these learners at the given time. The hazard function of the empirical data presents as a concave function, but it decreases in the tail. The shape of the concave segment is a bathtub curve[27], while the decreasing tail segment shows the Lindy effect[28]: the more you have learned, the more you want to learn. The features of the hazard function have the potential to help in assessing the attractiveness of courses.

We analyze the evolutionary mechanisms underlying these features. We categorize learners as lognormal learners and segment learners based on the different features of their viewing duration distribution, namely lognormal distribution and power law with exponential cutoff. For each category, we provide a random differential equation to describe the growth mechanism of its viewing durations. For lognormal learners, we assume that their duration change rate correlates to their past duration and to a random disturbance. For segment learners, we assume that their duration change rate is inversely proportional to time, and treat their starting time of viewing a course as random. The equations express the factor of memory in viewing behavior for lognormal learners, and that of memorylessness for segment learners. The solutions to the equations reproduce the features of the density function and those of the hazard function.

This paper is organized as follows. Empirical data are described in Section 2. The hazard function of viewing durations is described in Sections 3 – 6, where the evolutionary mechanisms of the hazard function are also analyzed. The results are discussed and conclusion is drawn in Section 7.

2 Data

We analyze the log data of eight courses from January 1, 2017 to November 10, 2017 on the iCourse platform. The courses are selected from the natural sciences, social sciences, humanities, and engineering and technology. Specific statistical indexes of these courses are listed in Table 1, which have been used to analyze course attractions in our previous work[29]. The data include the time length of each video. For each learner, the data include the viewing start time of each video he or she opened, and the corresponding viewing durations.

Videos can be downloaded with the iCourse app. The log data of viewing downloaded videos are also collected, unless the app disconnects from the Internet. Accordingly, our study only includes online viewing behavior, recorded as log data. However, learners may be off-line while a video is playing, in which case the viewing cannot be measured using log data; this is a limitation of our study. In addition, some typical video operations are not analyzed here, such as pausing, skipping, moving backwards or forwards, and changing speed.

We concentrate on learners' viewing durations, defined as the amount of time that a video is playing, without counting pauses. We introduce the following symbols to express the duration: supposing that learners $\{L_1, \ldots, L_m\}$ have viewed a course with n videos $\{V_1, \ldots, V_n\}$, we denote the time length of video V_i as l_i and the duration of learner L_s viewing V_i as t_i^s. The viewing duration of learner L_s on the course is then $\sum_{i=1}^{n} t_i^s$. Hereafter, the term "duration" is used as shorthand for the viewing duration on a course.

3 Survival Analysis of Viewing Durations

A learner's viewing duration on a course can be regarded as the "lifetime" of his or her viewing behavior. The number of learners with duration t expresses the number

Table 1 Specific statistical indexes of the data provided by iCourse. Index *m*: the number of learners; *n*: the number of videos; *a*: the number of all-rounders who viewed all of the videos; *b*: the average number of viewed videos per learner; *c*: the average viewing duration of learners (unit: hour); and *d*: the average time length of videos (unit: hour).

Course	Course ID	m	n	a	b	c	d
Calculus	1002301004	2955	129	2	8.081	0.998	0.189
Game theory	1002223009	4764	38	66	7.141	2.238	0.427
Finance	1002301014	6380	63	2	5.368	1.310	0.330
Psychology	1002301008	3827	26	59	5.008	0.913	0.204
Spoken English	1002299019	11 719	46	7	3.032	0.321	0.106
Etiquette	1002242007	3846	41	22	7.787	1.271	0.205
C Language	1002303013	17 541	81	39	12.47	1.541	0.142
Python	1002235009	13 417	53	28	10.32	0.896	0.087

of dropouts at "age" t, where $t \in [0, T]$, and T is the maximum viewing duration. Therefore, the density function of viewing durations, denoted as $f(t)$, expresses the rate of dropouts at any possible t. The rate of dropouts at a given t for the learners with a viewing duration of no less than t is calculated as $h(t) = f(t)/S(t)$, where $S(t) = \int_t^T f(\tau)\mathrm{d}\tau$ is the probability of a learner's duration being no less than t. In survival analysis[30], $S(t)$ and $h(t)$ are the survival function and hazard function, respectively. The function $h(t)$ is the derivative of $\log S(t)$, and thus is more informative about dropouts.

The hazard function of the empirical data presents as a concave function but with a decreasing tail (see Fig. 1). The shape of the concave segment is a bathtub curve, which is a concept taken from product quality assessment. Such a curve is often used to describe the failures of products over time, capturing the decreasing rate of early failures as defective products are discarded, followed by random failures at a constant rate during the useful life of products, and then an increasing rate of wear-out failures as the products exceed their designed lifetime. In this study, we call the dropouts on the increasing segment "wear-out dropouts". The decreasing tail is known as the Lindy effect, meaning that future life expectancy is proportional to current age. In the terms of our study, this means that every additional period of duration implies a longer remaining duration expectancy.

To understand why the Lindy effect and bathtub curve emerge simultaneously, we analyzed the features of the density function $f(t)$, and mined the dynamic mechanisms underneath those features, because the hazard function $h(t)$ is based on $f(t)$. When a learner has viewed a certain number of videos, his or her viewing duration follows a lognormal distribution (the results of the Kolmogorov-Smirnov (KS) test are shown in Table 2). For the remaining learners, most of their durations follow a power law with an exponential cutoff (the results of goodness-of-fit are given in Table 2). To illustrate these features, we fit the parameters of the distributions for each course, and generated synthetic durations following each distribution (see Fig. 2), the number of which is the same as that of the corresponding empirical durations. The comparison between empirical duration distributions and synthetic ones is shown in Fig. 3. What is the relationship between these features of the density function and the shape of the hazard function? To answer this question, we explore the evolutionary mechanisms underlying these features in the following sections.

4 Lindy Effect, Wear-Out Dropouts, and Lognormal Distribution

The empirical data show that the viewing durations of those learners who have viewed no less than τ videos follow a lognormal distribution (see Table 2). We call these lognormal learners, and discovered them using an

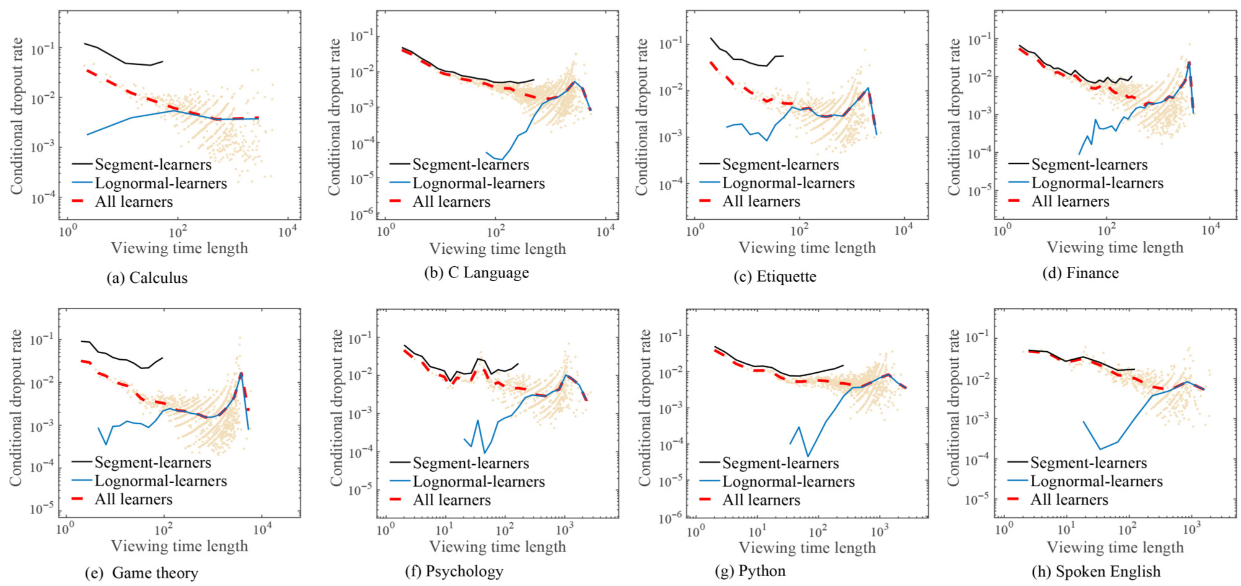


Fig. 1 Hazard functions of empirical viewing durations. Panels show the hazard functions (orange circles) and their trend line (red dotted lines) of viewing durations (unit = 2 seconds). Panels also show the trend lines of the hazard functions for lognormal learners (blue lines) and segment learners (black lines), respectively.

Table 2 Fitting parameters and goodness-of-fit.

Course	α	β	μ	σ	τ	p-value	ψ (%)
Calculus	0.7284	0.0035	6.1694	0.8155	11	0.059	15.40
C Language	0.5803	0.0028	6.9664	0.4865	29	0.067	12.48
Etiquette	0.5288	0.0151	5.5873	0.9395	4	0.069	16.09
Finance	0.7557	0.0024	6.6165	0.8052	8	0.092	14.53
Game theory	0.6556	0.0023	6.7109	0.7546	7	0.105	16.32
Psychology	0.5391	0.0062	6.0323	0.6659	7	0.355	23.05
Python	0.2622	0.0083	6.0817	0.4989	19	0.077	16.08
Spoken English	0.9561	0.0044	5.9030	0.5965	10	0.123	30.36

Note: The parameters of $x^{\alpha}e^{\beta x}$ are fitted through multiple linear regression. The parameters of lognormal(μ, σ) are calculated based on empirical data. At significance level 5%, the KS test cannot reject that the viewing durations of the learners, who have viewed no less than τ videos, follow a lognormal distribution (p-values > 0.05). The good-of-fit index ψ is the half of the cumulative difference between the duration distribution of segment-learners and the corresponding synthetic one.

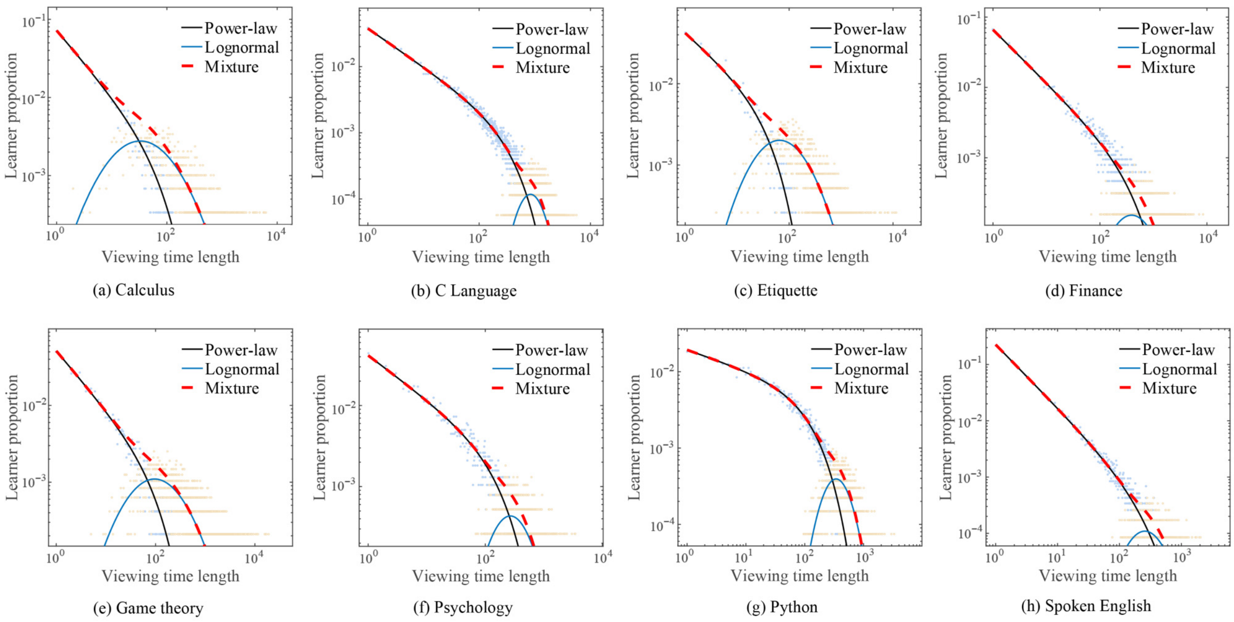


Fig. 2 Synthetic viewing duration distributions. Panels show the density functions of lognormal distributions (blue lines, orange circles) and those of the power-law distributions that have an exponential cutoff (black lines, blue circles) respectively, as well as the mixture density functions of them (red dotted lines). The parameters of these distributions are listed in Table 2.

algorithm given in Ref. [29] (see Table 3). Following an identical lognormal distribution means that this set of learners belongs to the same population in the sense of viewing duration; thus they can be categorized as a single class.

The density function of lognormal distribution is $f(x) = e^{\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}/(\sigma\sqrt{2\pi}x)$, where $x \in [1, \infty)$, $\sigma > 0$, and $\mu \in \mathbb{R}$. The corresponding hazard function is

$$h(x) = \frac{1}{x\sigma}\sqrt{\frac{2}{\pi}}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}\left(1-\text{erf}\left(\frac{\log x-\mu}{\sqrt{2}\sigma}\right)\right)^{-1} \quad (1)$$

The hazard function in Eq. (1) is convex when its corresponding density function is convex[31]. Figures 1 and 2 show the density function of lognormal learners and the corresponding hazard function to be convex for each course under study, revealing the wear-out dropouts and the Lindy effect. Figure 4 shows that the hazard functions of synthetic durations make a reasonable fit to those of the empirical data, which verifies the above arguments.

To find the evolutionary mechanisms underlying the wear-out dropouts and the Lindy effect, we return to where the notion of lognormal distributions emerged. They are often seen in the lifetime distributions of mechanical units[32], where the lifetime of a unit is affected by the multiplication of many small factors.

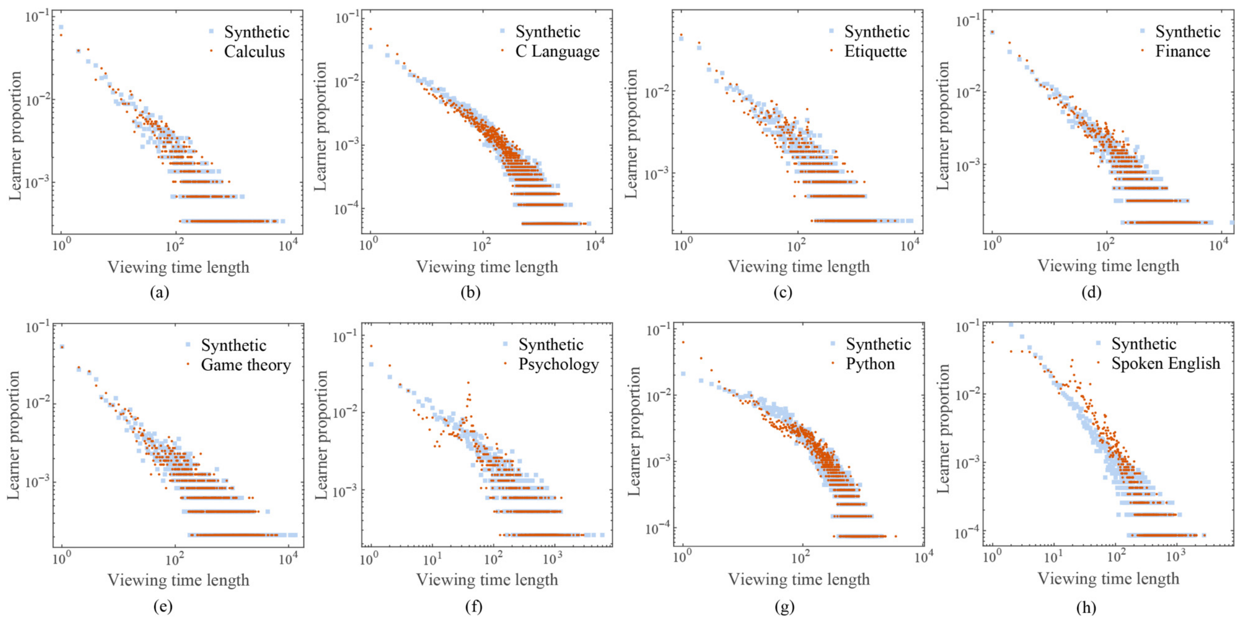


Fig. 3 Comparisons between the empirical distributions of viewing durations and synthetic ones. Panels show that each empirical distribution can be approximated by a mixture of a lognormal distribution and a power law with an exponential cutoff.

Table 3 An algorithm of categorizing learners[29].

Input: the viewing duration t_s and the number of viewed videos n_s of learners L_s $(s = 1, \ldots, m)$.
For k from 0 to $\max(n_1, \ldots, n_m)$: do the KS test for t_s of the learners L_s satisfying $n_s > k$ with the null hypothesis that they follow a lognormal distribution; break if the test cannot reject the null hypothesis at significance level 5%.
Output: the current k (denoted as κ).

Note: The unit of durations is millisecond. Categorize learner L_s as a lognormal-learner if $n_s > \kappa$, and as a segment-learner if else.

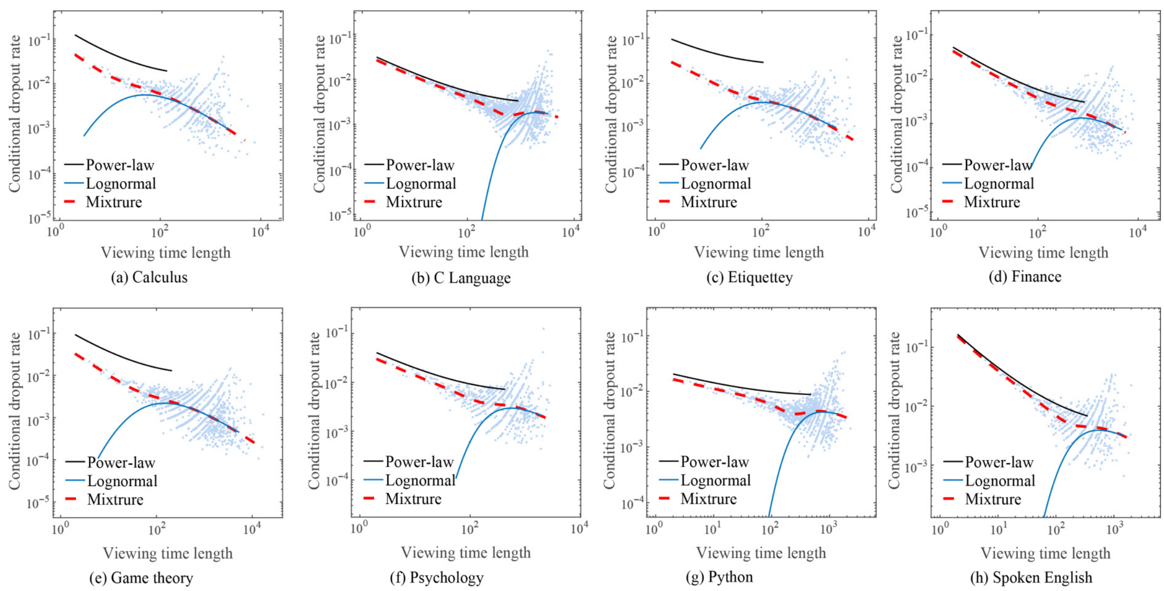


Fig. 4 Hazard functions of synthetic viewing durations. Panels show the hazard functions (blue squares) and their trend line (red dotted lines) of synthetic durations. Panels also show the hazard functions of lognormal distributions (blue lines) and those of the power-law distributions that have an exponential cutoff (black lines).

Approximating these factors as a range of independent and identically distributed random variables, the central limit theorem says that the summation of these variables in log scale follows a normal distribution. Transforming this to the original scale reveals that the multiplication of these factors follows a lognormal distribution. This is the multiplicative version of the central limit theorem, which is known as Gibrat's law[33].

The class of lognormal learners includes the "all-rounders": those who viewed all of the videos. The endurance of all-rounders and the remainder of the lognormal learners are therefore homogenous, such that lognormal learners could be regarded as potential all-rounders. For a potential all-rounder who is willing to complete a course, his endurance of viewing videos (measured by his viewing duration) could be analogized to a mechanical unit whose failure mode is of a fatigue-stress nature. The life of such a unit follows a lognormal distribution. This analogy enlightens us to provide:

$$\mathrm{d}k_i(t) = \mu k_i(t)\mathrm{d}t + \sigma k_i(t)\mathrm{d}w \tag{2}$$

where $k_i(t)$ is the duration at time t of learner i, w is the Wiener process, and μ and $\sigma > 0$. Supposing $k_i(t_0) = 1$ gives rise to the solution $k_i(t) = \mathrm{e}^{(\mu-\sigma^2/2)(t-t_0)+\sigma\int_{t_0}^{t}\mathrm{d}w}$, which is the random variable of a lognormal distribution.

Equation (2) entails that the viewing behavior of a lognormal learner has memory, because the change rate of the duration correlates to his or her past duration. Moreover, the expected change rate is positively correlated to the past duration. This means when t is large enough, the duration increases exponentially; the more you learn, the more you want to learn. This is the status of learners who are deeply impressed by a course.

5 Dropouts with a Constant Rate and Exponential Cutoff

Figure 3 shows the viewing durations of those learners who view less than τ videos (Table 2); the distribution is approximately a power law with an exponential cutoff. The emergence of the cutoff is mainly due to the durations of segment learners, which are no less than one minute, approximately following an exponential distribution (Fig. 5). The density function of the exponential distribution is $f(x) = \mathrm{e}^{-x/\lambda}/\lambda$, where $x \in [1, \infty)$, and $\lambda > 0$. The corresponding survivor function is $S(x) = \mathrm{e}^{-x/\lambda}$, and then the hazard function is a constant, $h(x) = f(x)/S(x) = 1/\lambda$.

To find the evolutionary mechanism underlying the dropouts with a constant rate, we return to the mechanism underlying exponential distribution. The distribution is characterized by memorylessness, because it satisfies $p(T > s + t | T > s) = \mathrm{e}^{-(s+t)/\lambda}/\mathrm{e}^{-s/\lambda} = \mathrm{e}^{-t/\lambda} = p(T > t)$ for any possible T, s, and t. In our study, memorylessness means that the future viewing duration is free from the influence of the past duration. For example, the probability of viewing a minute of video is the same for a learner who has previously viewed ten minutes of videos as for a learner

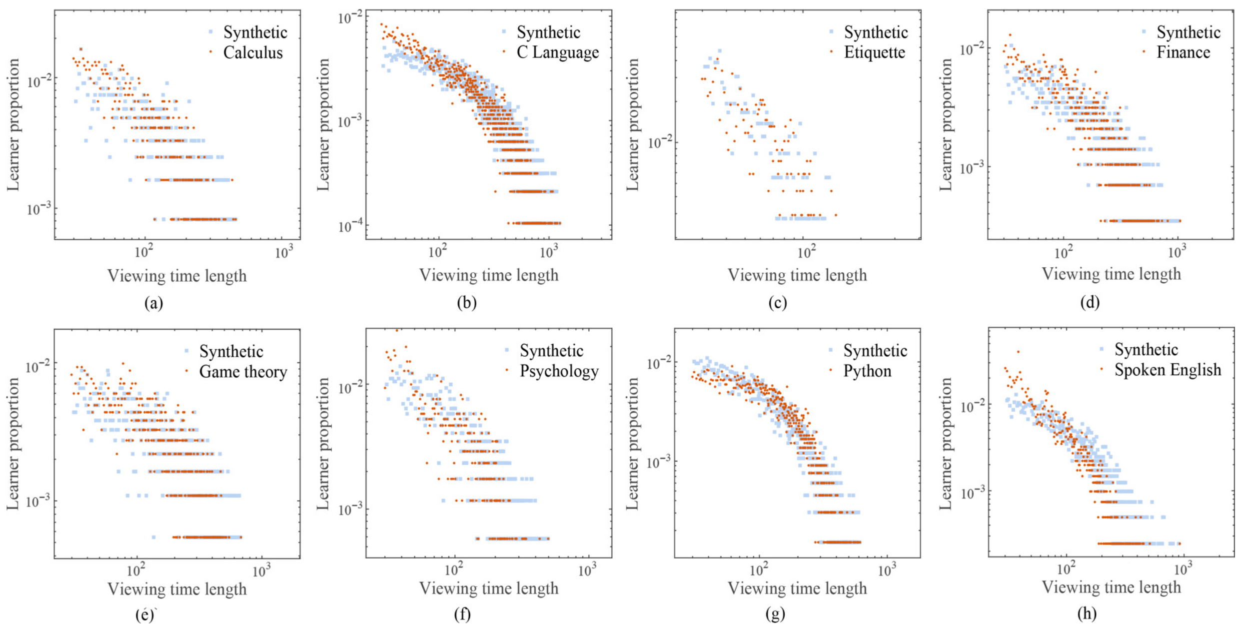


Fig. 5 Exponential cutoffs of viewing duration distributions. Panels show the duration distributions of the segment learners that viewed no less than one minute (red circles), compared with the predictions of Eq. (3) (blue squares).

who has not viewed any videos.

Due to memorylessness and learning fatigue, it is reasonable to suppose that the change rate of viewing duration will decrease with time. For simplicity, we assume that the change rate of learner i's viewing duration $k_i(t)$ would be

$$\frac{\mathrm{d}}{\mathrm{d}t}k_i(t) = \frac{\lambda}{t} \tag{3}$$

Solving the equation in the time interval $[y, T]$ gives rise to

$$k_i(T) = \lambda \log \frac{T}{y} \tag{4}$$

where y is the start time of viewing. Supposing that y is a random variable of the uniform distribution over $[T_0, T]$ gives rise to

$$p(k_i \leqslant x) = p\left(\lambda \log \frac{T}{y} \leqslant x\right) = p(T\mathrm{e}^{-\frac{x}{\lambda}} < y) = 1 - \frac{T}{T - T_0}\mathrm{e}^{-\frac{x}{\lambda}} \tag{5}$$

and leads to an exponential distribution:

$$p(k_i = x) = \frac{\mathrm{d}}{\mathrm{d}x}p(k_i \leqslant x) = \frac{T}{\lambda(T - T_0)}\mathrm{e}^{-\frac{x}{\lambda}} \tag{6}$$

When $T_0 = 0$, Eq. (6) is the standard exponential distribution.

To make the synthetic duration distributions fit the empirical ones, we valued the parameters of the solution in Eq. (4) based on the information from the empirical data. We set the domain of the durations of segment learners (which are no less than one minute) as $[T_1, T_2]$, and calculated the exponent $-1/\lambda$ of the formula in Eq. (6) by fitting the empirical data (see Table 4). Letting the simulated duration $\lambda \log(T/y)$ belong to $[T_1, T_2]$ gives the sampling interval $[\mathrm{e}^{-T_2/\lambda}, \mathrm{e}^{-T_1/\lambda}]$ for y/T. Table 5 shows the detail of this simulation process.

Analytical arguments allow for the prediction of the exponential cutoff. The simulations based on the solution in Eq. (4) also provide a reasonable fit to those of the empirical data (see Fig. 5). Therefore, Eq. (3) can be regarded as an expression of the evolutionary mechanism for the exponential cutoff and for the random dropouts with a constant rate.

6 Early Decreasing Dropouts and Power Law

The early decreasing trend of dropout rates appears in the hazard functions of the empirical data, describing a phenomenon in which the dropout rate of viewing a course decreases within the first minute. It describes the period of "infant mortality" during which learners who are only previewing a course stop viewing. Meanwhile, the density functions of the empirical data show that the viewing durations of segment learners who viewed less than one minute can be approximately fitted by a power-law function (see Fig. 6). Denoting the density function

Table 4 Parameters of synthetic power laws and exponential cutoffs.

Course	T	λ	N_1	ψ_1 (%)	a	b	c	N_2	ψ_2 (%)
Calculus	460	116.9	1217	26.38	1.30×10^{3}	3.682	7.25×10^{-2}	1169	11.89
C Language	1253	224.5	9587	16.27	8.30×10^{3}	2.383	7.21×10^{-2}	5587	8.5
Etiquette	206	42.62	683	22.36	18.9	2.122	7.71×10^{-2}	1079	10.29
Finance	1042	178.1	2875	24.59	3.98×10^{3}	4.093	6.98×10^{-2}	2448	7.52
Game theory	678	181.4	1834	25.08	1.47×10^{2}	2.904	7.51×10^{-2}	1408	7.74
Psychology	495	93.70	1718	31.49	10.4	2.170	5.54×10^{-2}	1188	17.51
Python	616	104.2	6607	14.59	2.70	1.351	6.94×10^{-2}	4261	13.49
Spoken English	916	93.70	4062	25.58	1.29×10^{36}	22.78	5.92×10^{-2}	7074	18.52

Note: Index T: the maximum duration of segment learners; λ: the parameter of Eq. (3); N_1 and ψ_1 (N_2 and ψ_2): the number of the segment learners with duration no less than (less than) one minute and the half of the cumulative difference between the duration distribution of those learners and the corresponding synthetic distribution; a and b: the parameters of Eq. (8); and c: the normalization coefficient of the formula in Eq. (7).

Table 5 Modelling exponential cutoffs.

Input: the durations ($\geqslant$ 1 minute) of segment-learners.
Regress the coefficients of $\delta\mathrm{e}^{-x/\lambda}$ based on the input.
Calculate the input's domain $[T_1, T_2]$.
For i in range 1 to the number of empirical durations:
sample a y/T from the uniform distribution over $[\mathrm{e}^{-T_2/\lambda}, \mathrm{e}^{-T_1/\lambda}]$;
substitute it into Eq. (4) to obtain a random integer;
append the integer to the list of synthetic durations.
Output: the list of synthetic durations.

Note: The unit of durations is 2 seconds.

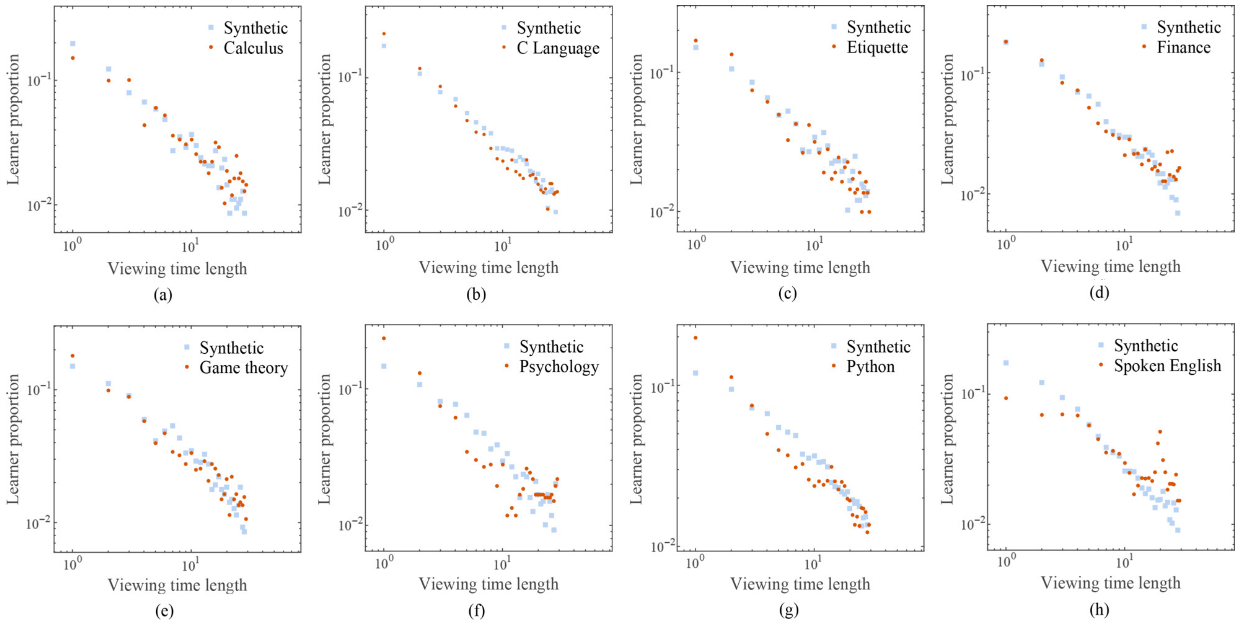


Fig. 6 Power-law parts of viewing duration distributions. Panels show the duration distributions of the segment learners who viewed less than one minute (red circles), compared with the predictions of Eq. (9) (blue squares).

of a power-law distribution as $f(x) = \beta x^{-\alpha}$, where β is the normalization coefficient, and $\alpha \in (0, 1)$, the random variable x is a value at a finite interval, denoted as $[R_1, R_2]$. Note that the value of the exponent α is different from that of degree distribution in network sciences, which is larger than one. The corresponding survivor function is $S(x) = 1 - \beta(x^{1-\alpha} - 1)/(1-\alpha)$, and the hazard function is

$$h(x) = \frac{\beta x^{-\alpha}}{1 - \dfrac{\beta}{1-\alpha}(x^{1-\alpha} - 1)} \tag{7}$$

where $h(x)$ is decreasing with the growth of x, when $x \leqslant R_2((\alpha+1)/2)^{1/(1-\alpha)}$.

To find the evolutionary mechanism underlying the early decreasing dropouts, we also return to the mechanism underlying such a power law. The durations of learners approximately following a power law are less than those approximately following an exponential distribution. Hence it is more reasonable to regard their learning behavior as memorylessness. Therefore, we suppose that the durations of those learners are also governed by Eq. (3). Meanwhile, the power law reflects the heterogeneity of samples[34–36], therefore the parameter λ in Eq. (3) should be heterogenous over learners.

We express this heterogeneity by $\lambda(\nu) = \nu^b/a$, and then

$$\frac{\mathrm{d}}{\mathrm{d}t} k_i(t) = \frac{\nu^b}{at} \tag{8}$$

where ν is a random integer of the uniform distribution over $[S_1, S_2]$, $a > 0$, and $b > 1$. Solving it on interval $[y, T]$ gives rise to

$$k_i(T) = \frac{\nu^b}{a} \log \frac{T}{y} \tag{9}$$

where y is the start time of viewing, sampled from the uniform distribution over $[T_0, T]$. Hence the expected value of $k_i(T)$ is $\dfrac{\nu^b \log 2}{a}$, which yields

$$p(k_i(T) \leqslant x) = p\left(\nu \leqslant \left(\frac{ax}{\log 2}\right)^{\frac{1}{b}}\right) = \frac{1}{(S_2 - S_1 + 1)} \left(\frac{ax}{\log 2}\right)^{\frac{1}{b}} \tag{10}$$

Then the density function of viewing duration is

$$p(k_i(T) = x) = \frac{\mathrm{d}}{\mathrm{d}x} p(k_i(T) \leqslant x) \propto x^{\frac{1}{b}-1} \tag{11}$$

The strict deduction of the density function needs averaging over all possible ν, which yields

$$p(k_i(T) = x) = \frac{1}{S_2 - S_1} \int_{S_1}^{S_2} \frac{1}{\lambda(\nu)} \mathrm{e}^{-\frac{1}{\lambda(\nu)}x} \mathrm{d}\nu = \frac{1}{S_2 - S_1} \int_{S_1}^{S_2} a\nu^{-b} \mathrm{e}^{-a\nu^{-b}x} \mathrm{d}\nu = \frac{a^{\frac{1}{b}}}{b(S_2 - S_1)} x^{\frac{1}{b}-1} \int_{aS_2^{-b}x}^{aS_1^{-b}x} \tau^{-\frac{1}{b}} \mathrm{e}^{-\tau} \mathrm{d}\tau \propto x^{\frac{1}{b}-1} I(x) \tag{12}$$

where $I(x) = \int_{aS_2^{-b}x}^{aS_1^{-b}x} \tau^{-\frac{1}{b}} \mathrm{e}^{-\tau} \mathrm{d}\tau$. Differentiating the integration part, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}I(x)=a^{1-\frac{1}{b}}x^{-\frac{1}{b}}\left(S_1^{1-b}\mathrm{e}^{-aS_1^{-b}x}-S_2^{1-b}\mathrm{e}^{-aS_2^{-b}x}\right) \tag{13}$$

This derivative is approximately equal to 0 if a is sufficiently large, which is guaranteed by the empirical values of a in Table 4. Hence the integration part is free of x and $p(k_i(T)=x)\propto x^{1/b-1}$.

To make the simulated distributions fit the empirical ones, we value the parameters of Eq. (9) based on the empirical data as follows. We calculate the domain of the durations of segment learners (which are less than one minute) $[R_1, R_2]$, and fit their distribution by the power law $cx^{-\alpha}$. The fitted values of α and c are listed in Tables 2 and 4, respectively. Comparing the coefficients of Eq. (11) to α and c gives rise to $\alpha = 1 - 1/b$ and $c = (a/\log 2)^{1/b}/((S_2 - S_1 + 1)b)$. Solving these obtains the value of a and b. The expected duration $\dfrac{\nu^b \log 2}{a}$ belonging to $[R_1, R_2]$ gives rise to the sampling interval for ν and then for y/T_2. Table 6 shows the detail for this simulation process.

The above analysis realizes a process of deriving the power law from a range of exponential distributions. Moreover, it provides an explanation for the early decreasing trend of the hazard functions. That is, the dropout rate $1/\lambda(\nu)$ decreases with the growth of the expected value $\lambda(\nu)$. The simulations based on the solution in Eq. (9) also provide a reasonable fit to the heads of the empirical duration distributions (see Fig. 5). Therefore, Eq. (8) can be regarded as an expression of the evolutionary mechanism for the power law and for the early decreasing trend. In addition, the memorylessness of Eq. (8) together with that of Eq. (3) can be regarded as the intrinsic meaning of the class of segment learners.

7 Discussion and Conclusion

The survival analysis on the viewing behavior of learners on MOOCs shows the hazard functions of empirical viewing durations are characterized by a bathtub curve and the Lindy effect simultaneously. Two random differential equations are provided to describe the growth processes of viewing durations. The solutions to these equations predict the features of the hazard functions. Therefore, these equations can be regarded as mathematical expressions of the evolutionary mechanisms underlying these features. The results are summarized in Fig. 7.

The results have the potential to illuminate specific aspects and implications in broader studies of learning behaviors. For example, the features of viewing duration distributions can be used to profile the type of learners, such as lognormal learners, those with a duration approximately following an exponential distribution, and those with a duration approximately following a power law. The fractions of these types vary over courses (see Fig. 8). Over half of the learners studying the course Calculus are lognormal learners, while almost half of the learners taking C Language or Python viewed less than one minute, with their duration approximately following a power law. Weighting each type with a different value helps to measure the attractiveness of MOOCs in a reasonable manner.

Comparing the duration distributions before and after adopting a teaching method can help to judge whether the method significantly increases or decreases learning durations. For example, if the KS test shows the duration distributions of lognormal learners to be identical before and after a change, this indicates that the adopted method has not made a significant improvement. This can also be used to compare the attractiveness of different courses; because it removes the heterogeneity of the number of learners, it gives a fair appraisal of courses with a high quality but few learners.

Table 6 Modelling power-law parts.

Input: the durations ($<$ 1 minute) of segment learners; the domain $[S_1, S_2]$ of ν.
Regress the coefficients of cx^{α} based on the input.
Calculate the parameters of Eq. (9): $b = 1/(1-\alpha)$, $a = (c(S_2 - S_1 + 1)b)^b \log 2$.
Calculate the input's domain $[R_1, R_2]$.
For i in range 1 to the number of empirical durations:
sample a ν from the uniform distribution over $[(aR_1/\log 2)^{1/b}, (aR_2/\log 2)^{1/b}]$;
sample a y/T_2 from the uniform distribution over $[\mathrm{e}^{-R_2/\lambda(\nu)}, \mathrm{e}^{-R_1/\lambda(\nu)}]$;
substitute them into Eq. (9) to obtain a random integer;
append the integer to the list of synthetic durations.
Output: the list of synthetic durations.

Note: The unit of durations is 2 seconds, $S_1 = 1$, and $S_2 = 29$.

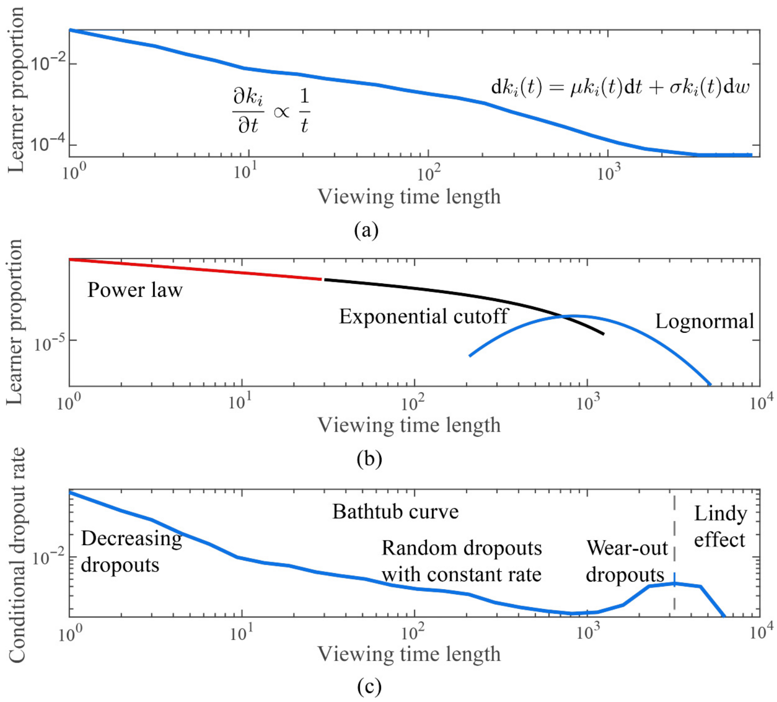


Fig. 7 An illustration of the presented results. The illustrated data are from the course C language. Panel (a) shows the density function and the evolutionary equations for viewing durations. Panels (b, c) show the features of the density function, and those of the corresponding hazard function. The equation on the left describes the evolutionary mechanism for power law with exponential cutoff, early deceasing dropouts, and the random dropouts with a constant rate. The equation on the right describes the mechanism for lognormal distribution, wear-out dropouts, and the Lindy effect.

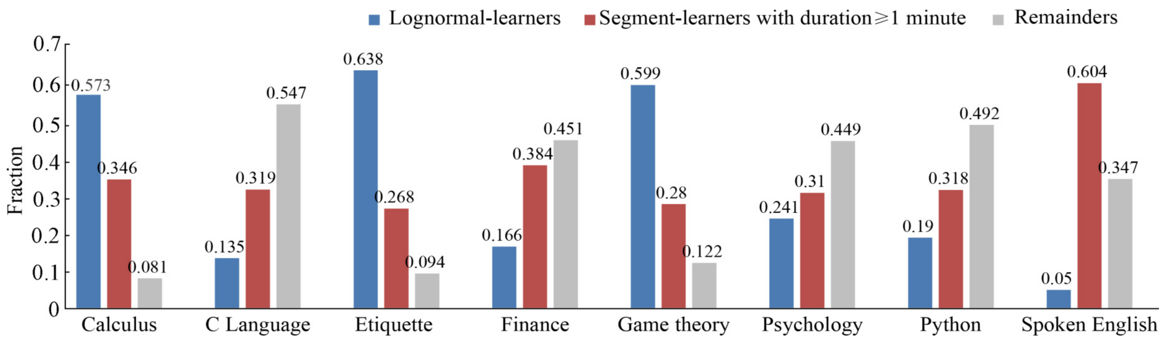


Fig. 8 Fractions of three learner types. The three types of learners are lognormal learners (the durations of them follow a lognormal distribution), the segment learners with duration no less than one minute (those durations follow an exponential distribution approximately), and the remainders (the durations of them follow a power law approximately).

In the presented equations, the viewing durations are based on random factors and memory or memorylessness. When they begin studying a course, a learner does not need to have knowledge of the course. As they study more deeply, however, the learner needs the knowledge from earlier in the course in order to proceed comfortably. This process could be regarded as the transition from memorylessness to memory. Meanwhile, the viewing duration distribution of each empirical course has a fat tail, known as a feature of complexity. We find that each tail is dominated by the tail of a lognormal distribution, and that viewing with memory can generate a lognormal distribution. Therefore, exploring the mechanisms underlying the transition contributes to an understanding of the role of memory in the complexity of learning behaviors.

Acknowledgment

The author thanks researcher Xiao Xiao at the Higher Education Press, Professor Ming Zhang at Peking University, and Professors Jinying Su and Jianping Li at the National University of Defense Technology for their helpful comments and feedback. The author is grateful to the iCourse MOOC platform for access to empirical data. This work was supported by the National Natural

Science Foundation of China (No. 61773020) and the National Education Science Foundation of China (No. DIA180383).

References

[1] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton, Studying learning in the worldwide classroom research into edX's first MOOC, *Res. Pract. Assess.*, vol. 8, pp. 13–25, 2013.

[2] M. N. Zhu, A. Sari, and M. M. Lee, A systematic review of research methods and topics of the empirical MOOC literature (2014–2016), *Int. High. Educ.*, vol. 37, pp. 31–39, 2018.

[3] E. J. Emanuel, Online education: MOOCs taken by educated few, *Nature*, vol. 503, no. 7476, p. 342, 2013.

[4] J. Reich, Rebooting MOOC research, *Science*, vol. 347, no. 6217, pp. 34–35, 2015.

[5] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, Engaging with massive online courses, in *Proc. 23rd Int. Conf. World Wide Web*, Seoul, Korea, 2014, pp. 687–698.

[6] K. Jona and S. Naidu, MOOCs: Emerging research, *Dist. Educ.*, vol. 35, no. 2, pp. 141–144, 2014.

[7] J. DeBoer, A. D. Ho, G. S. Stump, and L. Breslow, Changing "course": Reconceptualizing educational variables for massive open online courses, *Educ. Res.*, vol. 43, no. 2, pp. 74–84, 2014.

[8] J. P. Meyer and S. Zhu, Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating, *Res. Pract. Assess.*, vol. 8, pp. 26–39, 2013.

[9] A. S. Sunar, S. White, N. A. Abdullah, and H. C. Davis, How learners' interactions sustain engagement: A MOOC case study, *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 475–487, 2017.

[10] S. R. Emmons, R. P. Light, and K. Börner, MOOC visual analytics: Empowering students, teachers, researchers, and platform developers of massively open online courses, *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 10, pp. 2350–2363, 2017.

[11] C. Sandeen, Assessment's place in the new MOOC world, *Res. Pract. Assess.*, vol. 8, pp. 5–12, 2013.

[12] S. I. de Freitas, J. Morgan, and D. Gibson, Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision, *Br. J. Educ. Technol.*, vol. 46, no. 3, pp. 455–471, 2015.

[13] J. A. Greene, C. A. Oswald, and J. Pomerantz, Predictors of retention and achievement in a massive open online course, *Am. Educ. Res. J.*, vol. 52, no. 5, pp. 925–955, 2015.

[14] K. S. Hone and G. R. El Said, Exploring the factors affecting MOOC retention: A survey study, *Comput. Educ.*, vol. 98, pp. 157–168, 2016.

[15] A. Littlejohn, N. Hood, C. Milligan, and P. Mustain, Learning in MOOCs: Motivations and self-regulated learning in MOOCs, *Int. High. Educ.*, vol. 29, pp. 40–48, 2016.

[16] M. Barak, A. Watted, and H. Haick, Motivation to learn in massive open online courses: Examining aspects of language and social engagement, *Comput. Educ.*, vol. 94, pp. 49–60, 2016.

[17] P. G. de Barba, G. E. Kennedy, and M. D. Ainley, The role of students' motivation and participation in predicting performance in a MOOC, *J. Comput. Assist. Learn.*, vol. 32, no. 3, pp. 218–231, 2016.

[18] A. Watted and M. Barak, Motivating factors of MOOC completers: Comparing between university-affiliated students and general participants, *Int. High. Educ.*, vol. 37, pp. 11–20, 2018.

[19] S. J. Zheng, M. B. Rosson, P. C. Shih, and J. M. Carroll, Understanding student motivation, behaviors and perceptions in MOOCs, in *Proc. 18th ACM Conf. Computer Supported Cooperative Work & Social Computing*, Vancouver, Canada, 2015, pp. 1882–1895.

[20] K. M. Alraimi, H. Zo, and A. P. Ciganek, Understanding the MOOCs continuance: The role of openness and reputation, *Comput. Educ.*, vol. 80, pp. 28–38, 2015.

[21] K. Jordan, Initial trends in enrolment and completion of massive open online courses, *Int. Rev. Res. Open. Dist. Learn.*, vol. 15, no. 1, pp. 133–160, 2014.

[22] R. F. Kizilcec, C. Piech, and E. Schneider, Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses, in *Proc. 3rd Int. Conf. Learning Analytics and Knowledge*, Leuven, Belgium, 2013, pp. 170–179.

[23] K. F. Hew, Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCs, *Br. J. Educ. Technol.*, vol. 47, no. 2, pp. 320–341, 2016.

[24] L. Y. Li and C. C. Tsai, Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance, *Comput. Educ.*, vol. 114, pp. 286–297, 2017.

[25] G. Cheng and J. Chau, Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: An empirical study of a blended learning course, *Br. J. Educ. Technol.*, vol. 47, no. 2, pp. 257–278, 2016.

[26] C. R. Henrie, L. R. Halverson, and C. R. Graham, Measuring student engagement in technology-mediated learning: A review, *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[27] G. A. Klutke, P. C. Kiessler, and M. A. Wortman, A critical look at the bathtub curve, *IEEE Trans. Reliab.*, vol. 52, no. 1, pp. 125–129, 2003.

[28] J. Holman, Antifragile: Things that gain from disorder, *Quant. Finance*, vol. 13, no. 11, pp. 1691–1692, 2013.

[29] Z. Xie, Bridging MOOC education and information sciences: Empirical studies. *IEEE Access*, vol. 7, pp. 74206–74216, 2019.

[30] D. G. Kleinbaum and M. Klein, *Survival Analysis*. Springer, 2012.

[31] N. M. Kiefer, Economic duration data and hazard functions, *J. Econ. Lit.*, vol. 26, no. 2, pp. 646–679, 1988.

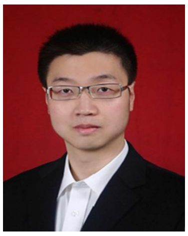
[32] J. H. Gaddum, Lognormal distributions, *Nature*, vol. 156, no. 3964, pp. 463–466, 1945.

[33] J. Sutton, Gibrat's legacy, *J. Econ. Lit.*, vol. 35, no. 1, pp. 40–59, 1997.

[34] Z. Xie, Z. L. Xie, M. Li, J. P. Li, and D. Y. Yi, Modeling the coevolution between citations and coauthorship of scientific papers, *Scientometrics*, vol. 112, no. 1, pp. 483–507, 2017.

[35] Z. Xie, Z. Z. Ouyang, J. P. Li, E. M. Dong, and D. Y. Yi, Modelling transition phenomena of scientific coauthorship networks, *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 2, pp. 305–317, 2018.

[36] Z. Xie, A cooperative game model for the multimodality of coauthorship networks, *Scientometrics*, doi: 10.1007/s11192-019-03183-z.



Zheng Xie received the PhD degree from Chinese Academy of Sciences in 2008. He was engaged in postdoctoral research at Zhejiang University between 2008 and 2009. He was a lecturer at National University of Defense Technology between 2010 and 2013. He was a visiting scholar at University of Bath between 2014 and 2015. He is now an associate professor with the Department of Mathematics at National University of Defense Technology. His research interests include geometric graph theory and its applications.