Collective behaviors of book holding durations

Ren-De Li [a,b], Qiang Guo [a], Jing-Ti Han [c], Jian-Guo Liu [a,c,*]

[a] *Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*
[b] *Library, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*
[c] *Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China*

ARTICLE INFO

ABSTRACT

Duration can directly reflect the collective reading behaviors of library user book holding. In this paper, by introducing the burstiness and memory coefficients, we empirically investigate the collective book holding behavior of three university libraries. The statistical results show that there are similar properties among the students with different backgrounds, presenting the burstiness $\langle B \rangle = -0.2$ and memory $\langle M \rangle = 0.5$ for three datasets, which indicates that memory and random effects coexist in student book holding durations. In addition, we analyze the behavior patterns without duplicate durations by merging a series of books borrowed and returned at the same time. The results show the average burstiness B increases to -0.16 and memory M drops to 0.16 for three datasets, which indicates that both duplicate behavior and student's preference affect the memory effect. Furthermore, we present a model which assumes student's next book holding duration follows the previous one with probability p, and with probability $1 - p$, the student would hold the book independently. The experimental results show that the presented model can reproduce the burstiness and memory effect of student book holding durations when $p = 0.5$ for empirical datasets and $p = 0.2$ for de-duplicate datasets, which indicate that the student's preferential holding behavior occurs with the probability p. This work helps in deeply understanding the regularity of duration-based human behaviors.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Collective behaviors of online users have been extensively investigated, which is of great significance for identifying the behavior patterns [1–3]. Oliveira et al. [4] found the scaling-law in Darwin's and Einstein's correspondence patterns. Brockman et al. [5] argued that the distribution of human travelling distances decayed as a power law. Saramäki [6] analyzed the mobile phone call pattern and found that human had persistence communication pattern. Besides the offline behaviors, the online behaviors, such as rating behaviors [7,8], posting behaviors [9–11], and web surfing behaviors [12–15], exhibited burstiness and memory effects. To explain the behavior patterns, the task- and interest-based models have been proposed. Task-based models [16–19] believed that human behaviors could be described as a decision-based process and tasks were executed according to their priorities. Interest-driven models [10,20–22] argued that the interest played an important role in human behaviors. Besides, circadian-driven models [23,24]

found that the day–night or weekly pattern existed in human behavior.

The book holding behavior may be affected by various endogenous and exogenous factors, including user's preference or interest [25], social influence [26], and the quality of book [27]. Vázquez et al. [19] investigated the inter-event time of borrowing behavior and found that it followed heavy-tailed distribution. To understand the inter-event time, the burstiness and memory coefficients were proposed [28]. Burstiness is the intermittent increases and decreases in activity or frequency of an event [1], measured by the coefficient of variation [28]. Memory is the similarity between two consecutive actions of an individual [7], calculated as a first-order autocorrelation function of time series [28]. Goh and Barabási [28] found that the inter-event time of the borrowing behavior had high-burstiness and low-memory properties. The above works mainly focus on the interval time between consecutive borrowing behaviors, which indicate the user's visiting pattern of book borrowing. However, book holding duration contains more information for understanding the user's reading behavior.

In this paper, we empirically investigate the collective behavior of the book holding durations. As illustrated in Fig. 1, the 1st, 2nd and 3rd books are borrowed at the same time but the hold-

---

* Corresponding author.
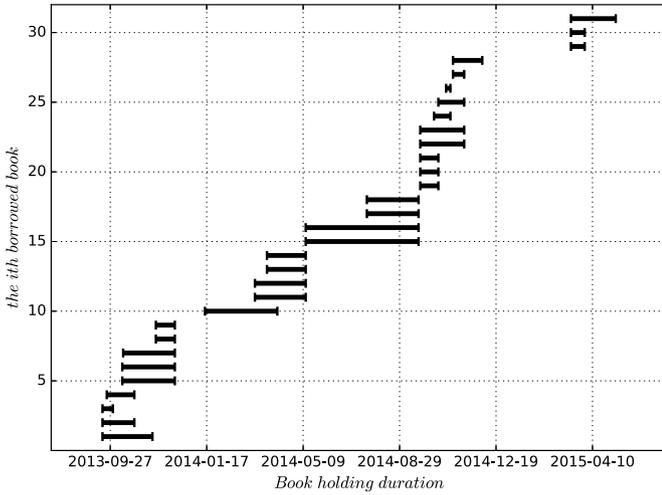*E-mail address:* liujg004@ustc.edu.cn (J.-G. Liu).

Fig. 1. The book holding durations of one student.

ing durations are different. The 10th and 11th books have similar holding durations although they are borrowed and returned in different time. And the 5th, 6th and 7th books are borrowed and returned at the same time, which are regarded as duplicate durations. The burstiness and memory effects are introduced to measure the collective properties. Then we present a model to reproduce the burstiness and memory effects of student book holding durations, and find that there are memory and random effects for the next book holding duration. Furthermore, we analyze the behavior patterns of three datasets without duplicate durations, and the results show that both duplicate behavior and student's preference affect the memory effect.

2. Empirical analysis

2.1. Data description

Three empirical datasets including the USST, SISU, and SNU are introduced in this paper which contain timestamps and student book borrowing-returning records (see Table 1 for basic statistics), where time span ranges from July 1, 2011 to June 30, 2015. The book holding duration from borrowing to returning is denoted by τ. The book holding duration sequence of student j is denoted as $\{\tau_{j1}, \tau_{j2}, \ldots, \tau_{jn_j}\}$, where n_j is the number of the borrowed books and $\tau_{j,i}$ is the holding duration of the ith book for student j. Since the collected data span four years, we use the average number of borrowed books per year $\langle n \rangle$ as average book borrowed index. In this paper, we only take students who had borrowed more than 10 books ($n_j > 10$) into consideration.

2.2. Measurements

For student j, burstiness B and memory M [28] are calculated to measure the physics of the student behaviors, in which the burstiness B_j of student j can be calculated by:

$$B_j \equiv \frac{(\sigma_{\tau_j}/m_{\tau_j} - 1)}{(\sigma_{\tau_j}/m_{\tau_j} + 1)} = \frac{\sigma_{\tau_j} - m_{\tau_j}}{\sigma_{\tau_j} + m_{\tau_j}}, \tag{1}$$

where m_{τ_j} and σ_{τ_j} are the mean and standard deviation of durations $\{\tau_{j1}, \tau_{j2}, \ldots, \tau_{jn_j}\}$. The value of burstiness B_j lies in $[-1, 1]$. The positive value of B_j means there are long/short book holding durations in regular durations and the negative value of B_j corresponds to regular durations for student j.

Table 1

Basic statistical properties of the datasets including the number of readers N_r, the number of books N_b, and the borrowing-returning records E for all students; the mode of book holding duration (unit: Day) $Mo(\tau)$, the mode of borrowed books per year $Mo(n)$ and the mode of standard deviation of book holding durations $Mo(\sigma_\tau)$ for the students who had borrowed more than 10 books.

Datasets	N_r	N_b	E	$Mo(\tau)$	$Mo(n)$	$Mo(\sigma_\tau)$
USST	29,988	222,606	651,548	31	11	25
SISU	14,392	147,164	411,831	33	11	24
SNU	52,282	365,676	955,449	29	11	16

Meanwhile, the memory M_j of student j is given as:

$$M_j \equiv \frac{1}{n_j - 1} \sum_{i=1}^{n_j-1} \frac{(\tau_{j,i} - m_1)(\tau_{j,i+1} - m_2)}{\sigma_1 \sigma_2}, \tag{2}$$

where $\tau_{j,i}$ is the holding duration of the ith book for student j, $m_1(m_2)$ and $\sigma_1(\sigma_2)$ are the mean and standard deviation of $\tau_{j,i}$ ($\tau_{j,i+1}$), respectively. The value of memory M_j lies in $[-1, 1]$. It's noted that M_j is positive when a long/short book holding duration tends to follow a long/short one, and M_j is negative when a short/long duration is likely to be followed by a long/short one for student j.

2.3. The numerical results

Fig. 2 shows the distribution of the burstiness B and memory M, from which one can find that the average burstiness $\langle B \rangle = -0.2$ and memory $\langle M \rangle = 0.5$ are the same in student book holding durations regardless of the student backgrounds.

After calculating burstiness B and memory M of each student's book holding durations, one can obtain the distributions of burstiness B and memory M for three empirical datasets. Then a null model is introduced to compare with the empirical results, which is constructed as follows: (i) Remain the borrowing-returning relations between students and books, (ii) Randomly generate each holding duration according to uniform distribution $U(0, 60)$, where the parameter 60 is set according to library policy: it is free for student to hold books for less than 60 days. In addition, if books are borrowed and returned at the same time, we merge these duplicate durations as one record, in order to eliminate the artificial enhancement of memory effect. The de-duplicate datasets are also used to compare with empirical results.

From Fig. 2, one can find that, for the null model, the average burstiness $\langle B \rangle = -0.28$ and memory $\langle M \rangle = 0$. The different results of burstiness between empirical and null model results suggest that empirical student book holding durations have random property. In addition, empirical memory $\langle M \rangle = 0.5$ is much larger than the value $\langle M \rangle = 0$ of the null model, which reports the strong memory effect on student book holding durations. Furthermore, the de-duplicate results show that the average burstiness B increases to -0.16 and memory M drops to 0.16 for three datasets comparing to the results of empirical datasets.

3. The model analysis

3.1. The model construction

To regenerate collective behaviors of book holding durations, we present a construction model. In this model, we assume that there are two different mechanisms for holding books, namely preferential book holding behavior and random book holding behavior. When a student reads a book, he/she might borrow other reference books with high probability of similar holding duration, which is denoted as preferential book holding behavior. Alternatively, when the student is interested in a new book, the holding
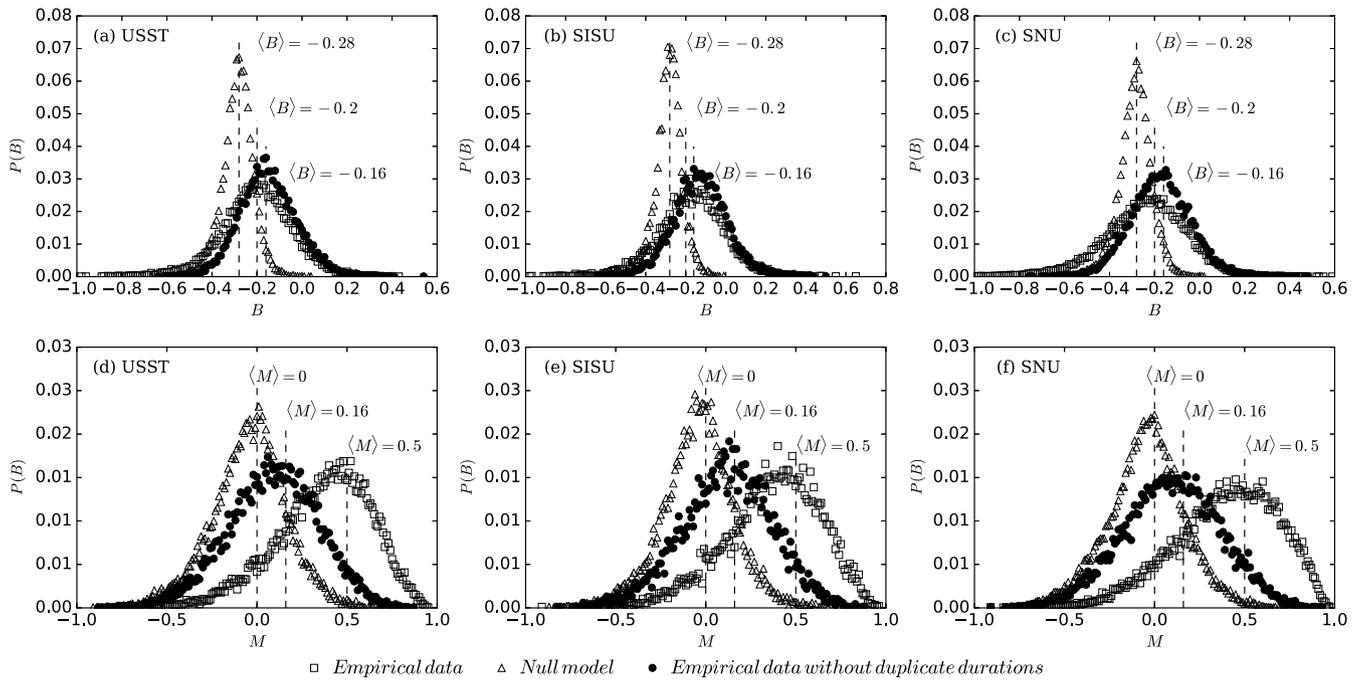
Fig. 2. The distributions of burstiness B and memory M for three empirical datasets which shows the similar distribution characteristics. The average burstiness $\langle B \rangle = -0.2$ is shown in subplots (a–c) and the memory $\langle M \rangle = 0.5$ is shown in subplots (e–f). The burstiness B and memory M of the null model are both normally distributed with the mean value $\langle B \rangle = -0.28$ and $\langle M \rangle = 0.5$, respectively. Empirical data without duplicate durations show the same burstiness $\langle B \rangle = -0.16$ and the memory $\langle M \rangle = 0.16$ for three datasets.

duration would be different, which is denoted as random book holding behavior. The model could be constructed as follows.

For each student j, there are two probability calculation methods of the next book holding duration $\tau_{j,i+1}$:

(i) Preferential book holding behavior, the next duration $\tau_{j,i+1}^a$ is generated by Gaussian distribution with probability p,

$$f(\tau_{j,i+1}^a) = \frac{1}{\sigma_\tau \sqrt{2\pi}} \exp\left(-\frac{(\tau_{j,i+1}^a - \tau_{j,i})^2}{2\sigma_\tau^2}\right), \tag{3}$$

where $\tau_{j,i}$ is the holding duration of ith book and σ_τ is the average standard deviation of book holding durations for all students, using $\sigma_\tau = 21$ as simulation value based on the average empirical parameters $Mo(\sigma_\tau)$ of three datasets. In this case, the generated time $\tau_{j,i+1}^a$ is similar to the previous one $\tau_{j,i}$. If $\tau_{j,i+1}^a$ falls in the interval of quantiles $[z_{p/2}, z_{1-p/2}]$ where $z_{p/2}$ is lower quantile and $z_{1-p/2}$ is upper quantile of Gaussian distribution, the next book holding duration $\tau_{j,i+1}$ will equal to the generated duration $|\tau_{j,i+1}^a|$. Otherwise, go to step (ii).

(ii) Duration $\tau_{j,i+1}^b$ is randomly generated by a uniform distribution with probability $1 - p$,

$$f(\tau_{j,i+1}^b) = \frac{1}{\tau_{\max} - \tau_{\min}}, \tag{4}$$

where $\tau_{\max} = 60$, which is selected according to library policy that lending period for books is normally 60 days, and $\tau_{\min} = 0$. In this case, the randomly generated time $\tau_{j,i+1}^b$ is independent of the previous one $\tau_{j,i}$. If this random mechanism is activated, the next book holding duration $\tau_{j,i+1}$ will equal to the generated duration $\tau_{j,i+1}^b$.

In this model, probability p is a free parameter. When regenerating the next book's holding duration, student book holding behavior obeys the preferential rule with probability p, or the random rule with probability $1 - p$. The larger the probability parameter p, the more preference the student's behavior tend to

have. Specially, when probability $p = 0$, the book holding duration is randomly regenerated.

3.2. The results analysis

In the simulation process, we remain the relations between the students and the borrowed books based on three data sets. For each student j, the number of borrowed books is simulated according to empirical number of borrowed books per year n_j. By setting a fixed value of probability parameter p, one can get durations of each student and then the burstiness and memory coefficients of book holding durations can be calculated. And by adjusting the value of the parameter p, one can reproduce different distributions of student book holding durations.

Fig. 3 shows the simulation results of construction model with $p = 0.5$. One can find that the burstiness distribution in the model shown in subplot (a–c) of Fig. 3 are similar to the results of empirical datasets. Specifically, for the USST, SISU, SUN datasets, the mean values of burstiness B in construction model and empirical datasets are both equal to -0.2. The memory effect can be found from subplot (d–f) of Fig. 3, in which average memory $M = 0.5$. The results indicate that student tends to preferential holding behavior with probability $p = 0.5$ and random holding behavior with probability $p = 0.5$.

In order to analyze the factor affecting the memory effect, we set probability $p = 0.2$ to compare results between de-duplicate datasets and construction model. As shown in Fig. 4, the construction model results of book holding duration are closely consistent with the de-duplicate results except the memory distribution in subplot (e). Specifically, as shown in subplot (d–f) of Fig. 4, the average memory of construction model and de-duplicate results are both $M = 0.16$. Similarly, the same results of burstiness can be found from subplot (a–c) of Fig. 4, in which average burstiness is -0.16. The results indicate that student tends to preferential holding behavior with probability $p = 0.2$ and random holding behavior with probability $p = 0.8$.
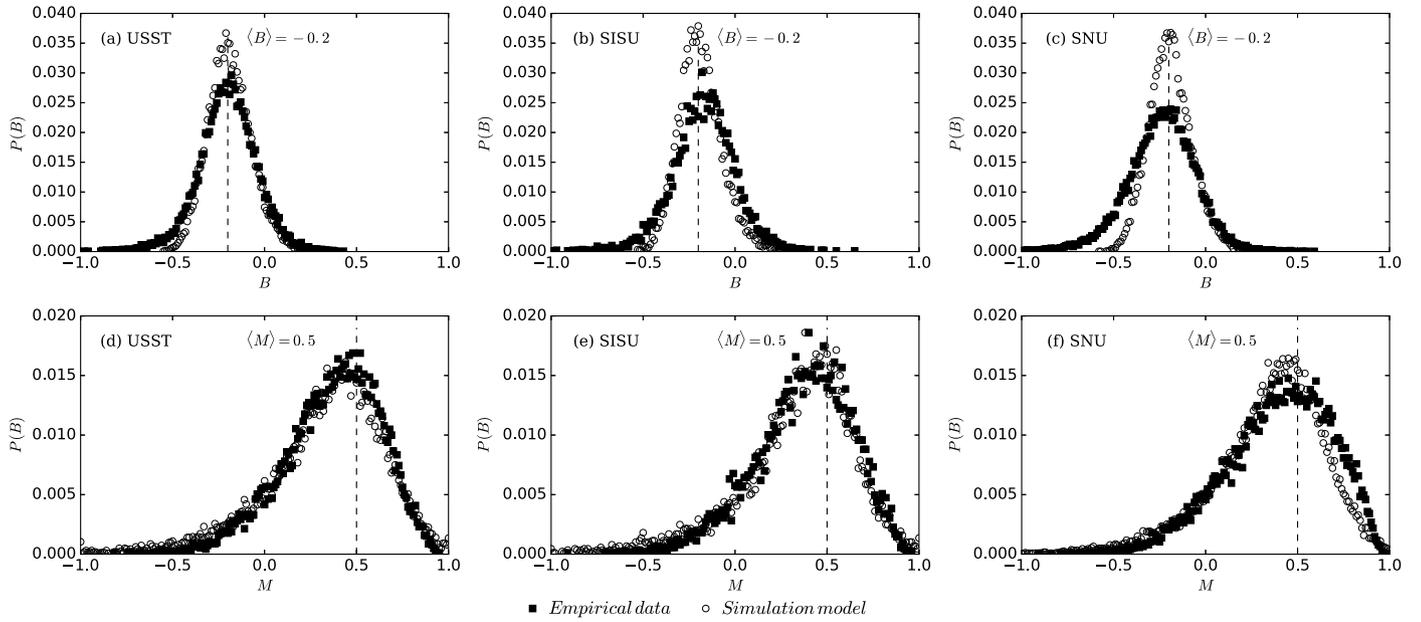
Fig. 3. The empirical distributions of burstiness B and memory M, comparing with simulation results when the parameter $p = 0.5$. The average burstiness $\langle B \rangle = -0.2$ is shown in subplots (a–c) and the memory $\langle M \rangle = 0.5$ is shown in subplots (e–f), each subplot from USST, SISU and SNU empirical and simulation results.
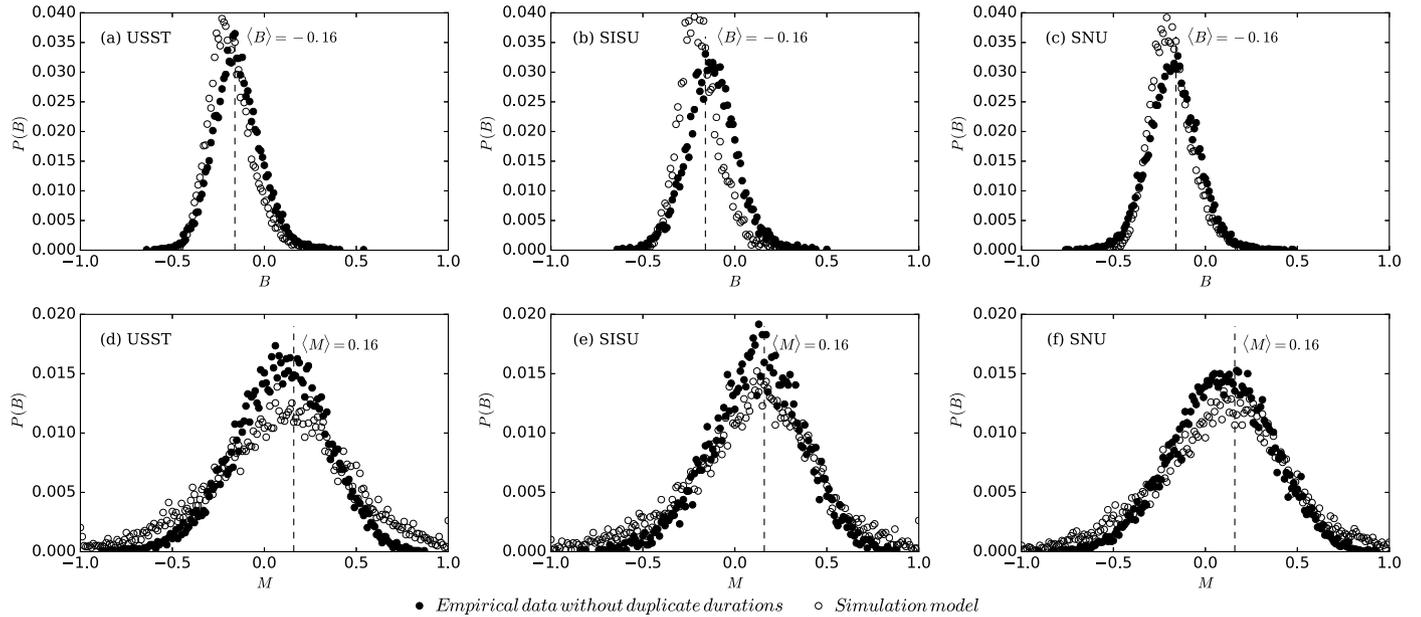

Fig. 4. The distributions of burstiness B and memory M for de-duplicate datasets, comparing with simulation results when the parameter $p = 0.2$. The average burstiness $\langle B \rangle = -0.16$ is shown in subplots (a–c) and the memory $\langle M \rangle = 0.16$ is shown in subplots (e–f) for USST, each subplot from SISU and SNU de-duplicate and simulation results.

Jensen–Shannon Divergence (short for JSD) [29] is introduced to measure the similarity between empirical distributions and construction model distributions of burstiness B and memory M. The JSD for two distributions X and Y can be expressed as,

$$JSD(X, Y) = \frac{1}{2} D(X \parallel Z) + \frac{1}{2} D(Y \parallel Z), \qquad (5)$$

where the distributions of parameters X and Y are compared, $Z = \frac{1}{2}(X + Y)$ is the distribution generated by the parameters X and Y, and $D(X \parallel Z)$ is the Kullback–Leibler divergence,

$$D(X \parallel Z) = \sum_i X(i) \log \frac{X(i)}{Z(i)}. \qquad (6)$$

Kullback–Leibler divergence $D(Y \parallel Z)$ is calculated in the same way. From the definition of JSD, one can find that the value of

Table 2
The value of Jensen–Shannon Divergence (JSD) between empirical and construction model distributions of the burstiness B and memory M.

Data sets	$p = 0.5$		$p = 0.2$	
	B	M	B	M
USST	0.0185	0.0113	0.0161	0.0346
SISU	0.0347	0.0121	0.0490	0.0267
SNU	0.0486	0.0115	0.0140	0.0200

JSD lies in $[-1, 1]$, and $JSD = 0$ only when two distributions are identical.

Table 2 shows the value of JSD between empirical results and construction model results of burstiness B and memory M, with the parameter $p = 0.5$ and $p = 0.2$. One can find that the values of JSD of each pair of distributions are less than 0.05 which

is close to 0. The results show that the distributions of burstiness B and memory M are close between construction model and empirical/de-duplicate datasets when adjusting the probability parameter p. Consequently, the results show that student's consecutive book holding behavior is depended on the previous one with probability 50% and random effect accounts for probability 50%. Furthermore, when removing the duplicate durations, student's consecutive book holding behavior depends on the previous one with probability 20% and random effect accounts for probability 80%.

4. Conclusion and discussions

In this paper, we empirically investigate student book borrowing-returning records for three universities, where the book holding duration is defined as the time between borrowing and returning. The statistical properties of student book holding durations are measured by burstiness coefficient B and memory coefficient M. Experimental results show that the average burstiness $\langle B \rangle = -0.2$ and memory $\langle M \rangle = 0.5$ for the three datasets, which suggest that there are memory and random effects in students' consecutive book holding durations regardless of their backgrounds. In addition, we analyze the behavior patterns of the datasets without duplicate durations. The results show that both duplicate behavior and student's preference affect the memory effect. When removing the duplicate effect, average burstiness B increases to -0.16 and memory M drops to 0.16. In order to regenerate the results of empirical datasets, we present a model by considering memory and random effects. The student might show preferential holding behavior with probability p, and appear random holding behavior with probability $1 - p$. By comparing the simulation results of book holding duration regenerated by construction model with empirical ones and de-duplicate datasets, one can find that: 1) If there are duplicate durations, student book holding behavior obeys the preferential rule with probability 50% or the random rule with probability 50% for the duration of the next book; 2) If duplicate durations are removed, student book holding behavior obeys the preferential rule with probability 20% and the random rule with probability 80% for the duration of the next book.

The comparison results of construction model with empirical datasets and de-duplicate datasets show that collective student book holding durations display memory and random effects, while both duplicate behavior and student's preference affect the memory effect. However, the proposed model considers the burstiness and memory separately. It is an open question to analyze the book holding duration patterns by considering burstiness and memory simultaneously as burstiness and memory may not be independent [30,31]. In addition, it is important to analyze the personal behavior patterns of book holding duration, which is of significance for recommendation [32,33].

References

[1] R. Lambiotte, L. Tabourier, J.C. Delvenne, Eur. Phys. J. B 86 (2013) 1.
[2] Q. Guo, W.J. Song, L. Hou, Y.L. Zhang, J.G. Liu, Physica A 401 (2014) 15.
[3] T. Zhou, X.P. Han, X.Y. Yan, Z.M. Yang, Z.D. Zhao, B.H. Wang, J. Univ. Electron. Sci. Tech. China 42 (2013) 481.
[4] J.G. Oliveira, A.L. Barabási, Nature 437 (2005) 1251.
[5] D. Brockman, L. Hufnagel, T. Geisel, Nature 439 (2006) 462.
[6] J. Saramäki, E.A. Leicht, E. López, S.G. Roberts, F. Reed-Tsochas, R.I. Dunbar, Proc. Natl. Acad. Sci. USA 111 (2014) 942.
[7] L. Hou, X. Pan, Q. Guo, J.G. Liu, Sci. Rep. 4 (2014) 6560.
[8] Y.L. Zhang, Q. Guo, J. Ni, J.G. Liu, Physica A 417 (2015) 261.
[9] P. Wang, T. Zhou, X.P. Han, B.H. Wang, Physica A 398 (2014) 145.
[10] Z.D. Zhao, Z.M. Yang, Z.K. Zhang, T. Zhou, Z.G. Huang, Y.C. Lai, Sci. Rep. 3 (2013) 3472.
[11] G.J. Ross, T. Jones, Phys. Rev. E 91 (2015) 062809.
[12] B. Gonçalves, J.J. Ramasco, Phys. Rev. E 78 (2008) 026123.
[13] F. Radicchi, Phys. Rev. E 80 (2009) 026118.
[14] J. Pan, H. Hu, Y. Liu, Physica A 413 (2014) 212.
[15] Z.D. Zhao, Z.G. Huang, L. Huang, H. Liu, Y.C. Lai, Phys. Rev. E 90 (2014) 050802.
[16] A.L. Barabási, Nature 435 (2005) 207.
[17] A. Vázquez, B. Racz, A. Lukacs, A.L. Barabási, Phys. Rev. Lett. 98 (2007) 158702.
[18] A. Vázquez, Phys. Rev. Lett. 95 (2005) 248701.
[19] A. Vázquez, J.G. Oliveira, Z. Dezsö, K.I. Goh, I. Kondor, A.L. Barabási, Phys. Rev. E 73 (2006) 036127.
[20] X.P. Han, T. Zhou, B.H. Wang, New J. Phys. 10 (2008) 073010.
[21] M.S. Shang, G.X. Chen, S.X. Dai, B.H. Wang, T. Zhou, Chin. Phys. Lett. 27 (2010) 048701.
[22] J.L. Guo, C. Fan, Z.H. Guo, Eur. Phys. J. B 81 (2011) 341.
[23] R.D. Malmgren, D.B. Stouffer, A.E. Motter, L.A. Amaral, Proc. Natl. Acad. Sci. USA 105 (2008) 18153.
[24] J. Kim, D. Lee, B. Kahng, PLoS ONE 3 (2013) e58292.
[25] L. Ji, J.G. Liu, L. Hou, Q. Guo, PLoS ONE 10 (2015) e0121105.
[26] J.G. Liu, Z.L. Hu, Q. Guo, Eur. Phys. J. B 86 (2013) 1.
[27] X.L. Liu, Q. Guo, L. Hou, C. Cheng, J.G. Liu, Physica A 436 (2015) 629.
[28] K.I. Goh, A.L. Barabási, Europhys. Lett. 81 (2008) 48002.
[29] J. Lin, IEEE Trans. Inf. Theory 37 (1991) 145.
[30] H.H. Jo, J.I. Perotti, K. Kaski, J. Kertsz, Phys. Rev. E 92 (2015) 022814.
[31] F.J. Guo, Z.M. Yang, Z.D. Zhao, T. Zhou, preprint, arXiv:1506.09096, 2015.
[32] J.G. Liu, T. Zhou, B.H. Wang, Y.C. Zhang, Q. Guo, Int. J. Mod. Phys. C 20 (2009) 1925.
[33] Q. Guo, F. Shao, Z.L. Hu, J.G. Liu, Europhys. Lett. 104 (2013) 28004.