Computers & Education 167 (2021) 104189

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

The learning behaviours of dropouts in MOOCs: A collective attention network perspective☆

Jingjing Zhang [a,*], Ming Gao [b], Jiang Zhang [c]

[a] Big Data Centre of Technology-mediated Education, Faculty of Education, Beijing Normal University, China
[b] Research Centre of Distance Education, Beijing Normal University, China
[c] School of Systems Science, Beijing Normal University, China

ARTICLE INFO

Keywords:
Distance education and online learning
Informal learning

ABSTRACT

Understanding the dropout phenomenon has advanced from viewing it as a sign of deficient quality to viewing it as an explicit sign of individual choice, which highlights the importance of investigating how dropouts learn in massive open online courses (MOOCs). Nevertheless, the short, limited and heterogeneous behaviours of individual dropouts create challenges for understanding how dropouts learn over time. Taking a systematic network perspective, this study used clickstream data to build a flow network model of collective attention to investigate how dropouts learn in XuetangX's Introduction to Psychology (2018 autumn). The results showed that the quantification of behavioural data presented a stereotypical image of dropouts, but the network analytics presented a rather different picture of how dropouts learn. Recognising the distinct roles of introductory learning resources could prevent dropping out and improve the accuracy of prediction models. Interestingly, the assessments embedded in the MOOCs performed a scaffolding role in guiding dropouts to learn. Thus, redesigning quizzes or examinations in open and flexible learning environments to construct a minimum cost network of collective attention is vital to making this online space cost effective for learners at risk.

1. Introduction

The increasing development of massive open online courses (MOOCs) is providing open and flexible learning experiences for a large body of learners worldwide (Adamopoulos, 2013). However, MOOCs' unusually high dropout rate—greater than 90% of participants—is alarming (Ang, Ge, & Seng, 2020; Jordan, 2014; Reich & Ruipérez-Valiente, 2019). Participants who complete a course are often regarded as persistent or successful learners, and their prevalence is frequently used to assess the quality of a course. As a result, high dropout rates are treated as a sign of deficient quality, which has yielded ample studies that identify the variables associated with dropout or completion rates (e.g., Gregori, Zhang, Galván-Fernández, & Fernández-Navarro, 2018; Pursel, Zhang, Jablokow, Choi, & Velegol, 2016; Romero-Rodriguez, Ramirez-Montoya, & Gonzalez, 2019). Currently, dropout studies are associated with personal

☆ Jingjing Zhang is Professor of Educational Technology at Beijing Normal University, where she directs the Big Data Centre for Technology-mediated Education. Ming Gao is currently a Ph.D. student at the Research Centre of Distance Education, Beijing Normal University. Jiang Zhang is a professor at School of Systems Science at Beijing Normal University. Address for correspondence: Jingjing Zhang, School of Educational Technology, Beijing Normal University, No. 19, XinJieKouWai Street, HaiDian District, Beijing, 100875, China.

* Corresponding author. Beijing Normal University, No. 19, XinJieKouWai Street, HaiDian District, Beijing, 100875, China.
E-mail address: jingjing.zhang@bnu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.compedu.2021.104189
Received 24 September 2020; Received in revised form 8 January 2021; Accepted 8 March 2021
Available online 17 March 2021
0360-1315/

intention and goals, design interventions and learning strategies, emphasising social aspects of learning and interactions in diverse academic settings and contexts. Refer to Lee and Choi (2011) and Aldowah, Al-Samarraie, Alzahrani, and Alalwan (2020) for a comprehensive review. Although we already know the significant predictors of and their correlations with dropping out, we often fail to provide effective interventions to increase retention in MOOCs, as too many factors have been identified in various MOOCs across diverse settings, adding to the difficulty of effectively guiding pedagogically sound intervention designs. Furthermore, carefully selected interventions that consider combined factors are too costly (e.g., adding to the instructors' workload) or impossible to apply to a larger population in different contexts.

Another major research trend considers dropping out to be an explicit sign of individual choice instead of deficient quality. The alarming dropout phenomenon is the result of the inherent nature of MOOCs in allowing individuals to choose and self-control their studies (Tschofen & Mackness, 2012). This counterargument highlights the importance of focusing on how dropouts learn in MOOCs and how the learning design of MOOCs facilitates their learning. Nevertheless, attempts to understand the needs, intentions or motivations of all MOOC learners are optimistic approaches to gaining a detailed picture of MOOC attrition but are often conducted on a small scale. In an open and flexible learning environment, targeting learners who are most likely to drop out becomes challenging, as such a large group of heterogeneous learners presents different profiles, including intrinsic and extrinsic characteristics and intentions (Hew, 2016).

Few studies have taken the learning analytics approach to understanding how dropouts learn in MOOCs, as it seems that few or no data are available to meaningfully use the data-mining approach. Furthermore, the behavioural data of dropouts have often been removed in the earlier learning analytics studies, as traditional statistical methods generally consider these data 'outliers'. Even in social network analyses, which are often performed to examine learning at scale, dropout behaviours are often regarded as 'insignificant' interactions, with the assumption that these "noisy" data need to be cleaned to reveal the real picture of online interaction. Additionally, classical social network modelling fails to account for steeply unequal MOOC participation patterns due to dropouts joining and leaving in a flexible manner (Zhang, Lou, Zhang, & Zhang, 2019). A social network that uses nodes to represent learners who might drop out is not a robust model for capturing the dynamics of learning interactions; consequently, the interpretation of such a model might be biased. The field of learning analytics tends to use a small fraction of the massive dataset of learners to yield a new understanding of learning at scale but has for years, ironically, overlooked the remaining learning population and the way they learn. To address this problem, we build upon earlier studies by using the network model of collective attention to investigate how dropouts learn at the collective level in a MOOC. A key innovation is the focus on using a new method to model behaviours as attention flow by considering dropouts—whose behavioural trajectories are short, limited and heterogeneous—to understand their learning patterns.

1.1. Literature review

High dropout rates have become the greatest challenge for MOOCs (Brown, Costello, Donlon, & Giolla-Mhichil, 2015; Gregori et al., 2018; Littlejohn, Hood, Milligan, & Mustain, 2016). In higher education, the definition of a dropout dates back to the work of Tinto (1975), who defines college dropouts in 'Dropout from higher education: A theoretical synthesis of recent research' as students who leave without receiving an end qualification and students who attend one or more educational institutions but never receive an end qualification from any of them. There is no official definition of dropouts in the MOOC literature. Earlier studies treat as dropouts students who attend a course but leave without attending the final examination or earning the course certification (Breslow et al., 2013; Jordan, 2014). Later, scholars tentatively refer to dropouts as different cohorts of students in different settings, e.g., some consider dropouts as students who do not submit final assignments or those who do not engage in the final week, while some consider dropouts as those who are inactive for up to two weeks (Sunar, White, Abdullah, & Davis, 2017) or those who do not complete the units they intended to (Henderikx, Kreijns, & Kalz, 2017). The reasons dropouts have been defined in different ways reflect upon the fact that researchers have continuously battled with high attrition in MOOCs for different purposes, for example, providing pedagogically sound interventions, learning strategies, innovative techniques or motivational methods to support dropouts. Dropout research has increased since 2014 and it has consistently been an active research area in the past 6 years (Rasheed, Kamsin, Abdullah, Zakari, & Haruna, 2019). Nevertheless, completion rates have not yet improved (Reich & Ruipérez-Valiente, 2019), which further highlights the complexity of this well-researched but challenging problem.

1.1.1. Critical view of dropouts—deficient quality and unsuccessful learners

Participants who cannot complete a course are often regarded as unsuccessful learners, which has yielded ample studies investigating factors associated with dropout or completion rates. Many efforts have been made to investigate learners' profiles using either self-reported questionnaires or interviews, such as demographic characteristics (van de Oudeweetering & Agirdag, 2018), self-efficacy (Abeer & Miri, 2014), emotion (Hillaire, Iniesto, & Rienties, 2017), personalisation (Abe, 2020), commitment, attitudes or motivation (Barak, Watted, & Haick, 2016; Kizilcec & Halawa, 2015; Pursel et al., 2016; Shapiro et al., 2017; Terras & Ramsay, 2015; Watted & Barak, 2018), competence, readiness, or prior experience (Al-Adwan & Khdour, 2020; Breslow et al., 2013; Greene, Oswald, & Pomerantz, 2015). These studies have pointed out that MOOC learners who are likely to drop out share similar demographic characteristics and personal factors with students who are likely to be unsuccessful in conventional education. The rationale for exploring these issues was to provide early interventions for learners who share similar profiles (e.g., Li et al., 2016). Nevertheless, there are far fewer people participate in the studies mentioned above compared with those who participate in MOOCs. This problem raises a major concern in MOOC studies that attempt to use self-reported methods, as the majority of these studies are limited to the fraction of learners who are willing to report from their own perspectives.

Another group of studies has attempted to identify factors contributing to high attrition rates, which are associated with course logistics, design and learning strategies in varied courses and settings, as a group of researchers realise that effective pedagogical interventions could further increase interaction, commitment, and ultimately engagement (Barak et al., 2016; Terras & Ramsay, 2015; Watted & Barak, 2018). They assume that MOOCs that are designed with deficient quality result in high dropout rates. As course-related factors are the researchers' anticipated reasons for the dropouts, the researchers tend to use large-scale studies to include factors that predict dropouts. For example, Jordan (2014) conducted over 39 courses with a wide sample of MOOC users and identified that course lengths and designs as potential factors influencing dropouts. Similarly, more studies have reported that the duration of course activity is the core factor influencing high dropout rates (Jiang, Williams, Schenke, Warschauer, & Dowd, 2014; Kloft, Stiehler, Zheng, & Pinkwart, 2014). In addition to course length, course design factors were explored using teachers' presence (Hone & Said, 2016) and accessibility (Hew, 2016), assessment and feedback (Bonk et al., 2018; Olivé, Huynh, Reynolds, Dougiamas, & Wiese, 2020), learner support (Gregori et al., 2018), and gamification (Romero-Rodriguez et al., 2019). In addition, a group of researchers (e.g., Rosé et al., 2014; Wang, Guo, He, & Wu, 2019; Yang, Wen, & Rosé, 2014) attributed high dropout rates to interactions with instructors and peers, which highlights the importance of the social aspect of MOOC learning. The rationale of these studies is usually to suggest pedagogically sound interventions or innovative learning strategies. Nevertheless, although we know the significant predictors and their correlations with dropping out, we often fail to provide effective interventions to increase retention in MOOCs, as too many factors have been identified in varied MOOCs across diverse settings, which adds to the difficulty of effectively guiding pedagogically sound intervention design. Furthermore, carefully selected interventions that consider combined factors are too costly (e.g., adding to the instructors' workload) or aree impossible to apply to a larger population in different contexts (e.g., gamification).

As we can see, dropout research tends to consider all of the attributes dropouts might have for prediction, with the assumption that dropouts' attributes and behaviours imply that they are unsuccessful learners or that course designs or learning strategies fail them. Of course, there might be many other reasons for not completing the course, including personal (62.5%), circumstantial/social (50%), course (47.5%), and academic (42.5%) factors, reported by Aldowah et al. (2020), echoing previous studies (e.g., Bonk et al., 2018). Refer to Lee and Choi (2011) and Aldowah et al. (2020) for a comprehensive review.

1.1.2. *Alternative explanation for dropping out—Accommodating individual choice*

Recently, our understanding of the dropping out phenomenon has progressed from being a sign of deficient quality to being an explicit sign of individual choice. Contributing to high attrition rates are diverse personal goals and interests (Dai, Teo, Rappa, & Huang, 2020; Henderikx et al., 2017), which are recognised as the consequences of MOOCs' inherent natural qualities of openness, massiveness and flexibility. Driven by different learner intentions, some dropouts may not be unsuccessful learners (e.g., Henderikx et al., 2017; Romero-Rodriguez et al., 2019; Zheng, Rosson, Shih, & Carroll, 2015).

Considering learners' diverse learning goals, Henderikx, Kreijins and Kalz (2017) proposed redefining dropout as learners' not completing the units they intended to; however, the dropout rate remains substantially high (approximately 30–41%) in two selected MOOCs. Learners have many reasons for not completing the course, and they present vastly different patterns of behaviours. Some of them never accessed Courseware, some randomly navigated certain pages, while others stopped participating after one or two weeks (Yang, Sinha, Adamson, & Rose, 2013). Those who intended to complete the course but failed to continue were struggling learners who certainly merit more support. This points more to the fact that MOOC attrition continues to be a major issue that merits rigorous studies investigating thee varied behaviours of dropouts (Ang et al., 2020).

To identify at-risk or struggling students in the learning process, a group of researchers performed a contextual analysis on forum data to detect confusion (Yang, Wen, Howley, Kraut, & Rosé, 2015), motivation (Wen, Yang, & Rosé, 2014), and sentiment associated with disengagement (Ramesh, Goldwasser, Huang, Daum, & Getoor, 2013). Nevertheless, forum interactions might not provide a meaningful context for studying dropouts, given that not all at-risk students continually participate in MOOC forums. A MOOC forum presents a rich-get-richer effect (Zhang, Skryabin, & Song, 2016). Previous studies (e.g., Tang, Xing, & Pei, 2018) have shown that only a small portion of learners posted on the forum, which points to the importance of using clickstream data to identify dropouts' varied patterns of activities. Additionally, given that MOOCs accommodate learning at scale, effective learning strategies (e.g., instructors' efforts to create and maintain an interactive discussion forum) that could operate well in small-scale scenarios are unlikely to work well in the MOOC environment. Of course, many of these pedagogically sound interventions could be adopted in small private online courses, which are characterised as having significant learner-instructor and learner-learner interactions among a small number of learners facilitated by instructors (Filius et al., 2018). Nevertheless, the current format of MOOCs can still signify the possibility of accommodating open access to learning at scale, which has important value for many people in different contexts and countries. The feweer regulation MOOCs defined in such an open and flexible learning environment, the greater challenges that are presented to instructors as they need to identify new ways of holding learners' attention (Garreta-Domingo, Sloep, & Hernández-Leo, 2018). Taking a learner-centred approach, the instructors role shifts from taking care of each individual learner to shaping the design of the course structure and guiding online learning based on knowledge about how learners behave or learn in MOOCs.

1.2. *A systematic view to understand collective patterns from the diverse learning behaviours of dropouts*

The diverse approaches learners take in learning online are highly individualised. Nevertheless, being able to understand dropout behaviours over time that are heterogeneous is an important prerequisite for effective learning design (Chen, Sonnert, Sadler, Sasselov, & Fredericks, 2020). The significance of using learning analytics approaches and methods to understand how dropouts behave online has been well documented (e.g., Aldowah et al., 2020; Ferguson & Clow, 2015; M.; Khalil & Ebner, 2017; Sunar et al., 2017). Many MOOC learners exhibit reading and observing behaviours rather than creating content, editing material, or interacting with peers or instructors (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014; Sun, Rau, & Ma, 2014).

The complexity of the dynamic nature, varied setting and contexts, and individual learning choices and intentions creates difficulties in understanding the learning behaviours of dropouts at the individual level. Furthermore, as attrition occurs over time, understanding how dropouts' behaviours change online is essential but often difficult (Yang et al., 2015). Specifically, as Garreta-Domingo et al. (2018) argued, the innate complexity of the online and flexible environment (which lacks regulations) highlights the importance and challenge of exploring emerging behavioural patterns we have not traditionally observed for dropouts. More recently, researchers who made use of clickstream to assess behaviours online also stressed the centrality of understanding behaviours from a network perspective (see Poquet, Hecking, & Chen, 2020). The concept of collective attention, taking an ecological perspective, conceptualises clickstream data as an embodiment of a continuous attention flow (Kammenhuber, Luxenburger, Feldmann, & Weikum, 2006), which has been proposed to account for this difficulty in exploring online behaviours (Wu & Huberman, 2007). It creates an open-flow model that highlights the capability of allowing dropouts to take short and limited learning trajectories that have often been dismissed but attempts to make sense of these behaviours at the collective level, including online news reading behaviours (Li & Tang, 2020), knowledge creation (Ciampaglia, Flammini, & Menczer, 2015), stock fluctuations (Liu, Yang, Li, & Yu, 2018), and human behaviour in elections (Eom, Puliga, Smailovic, Mozetic, & Caldarelli, 2015).

Despite recognising the observed diverse learning behaviours of dropouts, research using short, limited and heterogeneous behaviours has been limited, in particular, research involving the analysis of behaviours at short time intervals. To date, research investigating dropout behaviour has been carried out using learning behaviours for the whole course, not for each time period dropouts go online. For example, Chen et al.'s recent work (2020) examined dropouts' behaviours at chapter transitions and discovered that learners are more likely to drop out at chapter closures. The analysis of dropout behaviour for the period of the course presents stereotypical images of dropouts; that is, the quantity and frequency of learning behaviours decrease as the course progresses (Chen, Sonnert, Sadler, Sasselov, & Fredericks, 2020). Their study highlights the importance of using learning behaviours to inform interventions from the perspective of how dropouts learn, which has not previously been explored. Furthermore, using the time period when dropouts go online as an entity resolution, even the short, limited and heterogeneous behaviours of dropouts within each time period (learning session) can be meaningfully addressed using more advanced learning analytics techniques. For example, Zhang et al. (2019) and Zeng, Zhang, Gao, Xu, and Zhang (2020) examined learning behaviours at short time intervals taking the ecological perspective, which shed light on how learners interact with the structure of learning resources in depth and how learning behaviours as

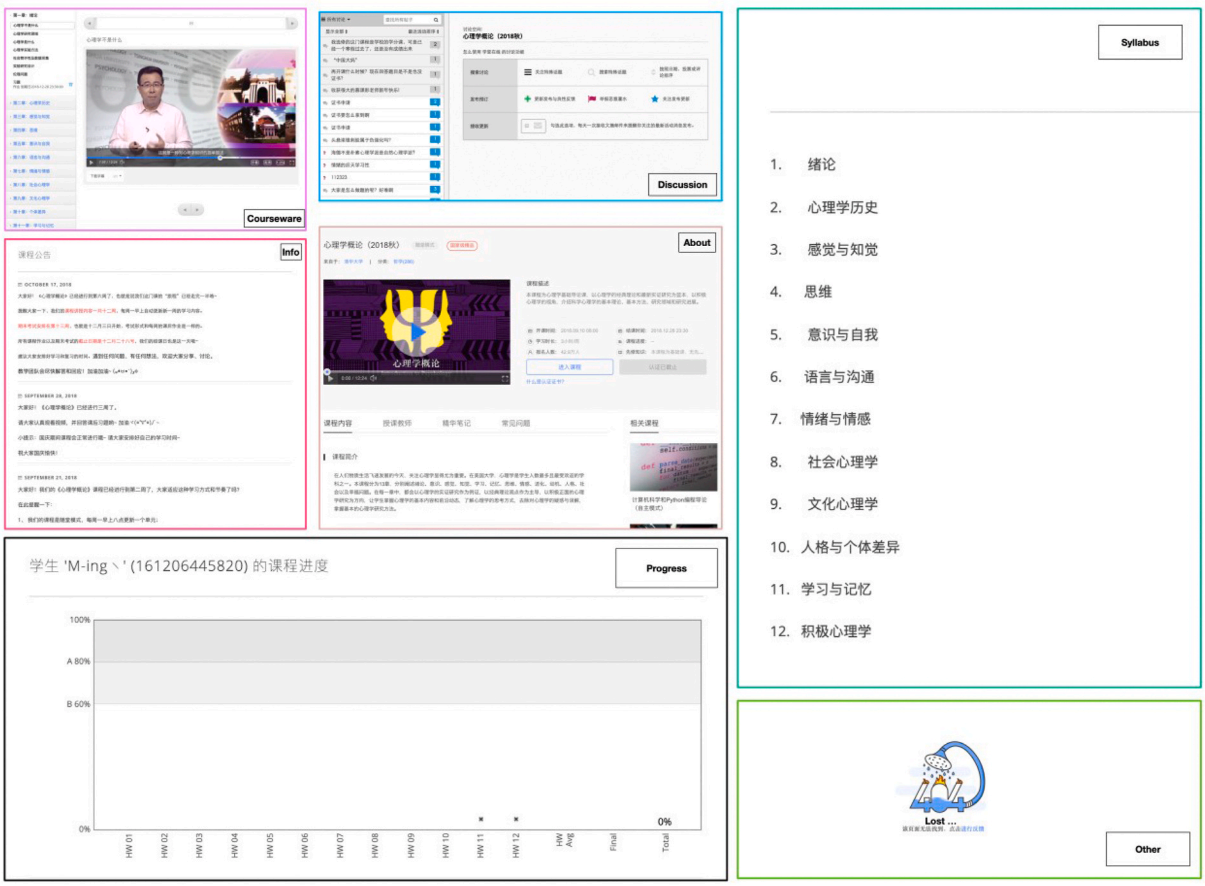

Fig. 1. Screenshot of introduction to psychology on XuetangX.

collective attention flows change as the course progresses. As the ecological view of online behaviours considers all of the actions dropouts take—including watching videos, accessing the course syllabus, and interacting with the forum—it understands dropouts' behaviour as inextricably woven into the varied levels of the complexity of learning.

2. Methods

2.1. Context

XuetangX is the first Chinese MOOC platform, which was initiated by Tsinghua University and the Ministry of Education Research Centre for Online Education (http://www.xuetangx.com/global). XuetangX's Introduction to Psychology (2018 autumn) course was selected as the case for this study. It was one of the first courses offered at XuetangX, and, since then, more than 570,000 learners have participated in the course. This course was selected as one of the 490 National Top-Quality Open Courses in 2018 (Ministry of Education of the People's Republic of China, 2018). In another study (Zhang & Yang, 2019), similar behavioural patterns in terms of quantity and frequency were presented in the five rounds of this course (2015–2017). As the overall instructional design of this course has remained the same, 2018 autumn was selected for the case study. This course offers 70 learning resources—including videos, quizzes and an examination—within 13 units (see Fig. 1).

A total of 9508 learners participated in the course; their behavioural data were automatically stored in the database. Notably, with regard to our previous studies, which involved Introduction to Psychology 2015 and 2016, more learners have tended to use mobile devices to access this course over time. Unfortunately, behavioural data created via mobile devices were not stored in a complete form, and, as a result, this study employed the behavioural data stored via the web for analysis, which includes approximately half of the learners (4271). Of these participants, 2110 lacked either registration information or examination records. Only 2161 of the participants had complete records in the database, among whom 33 participants passed the examination, 236 participants failed it, and 1892 participants did not take the exam. Among these 1892 learners, 14 began the examination but did not complete it, and the remaining learners did not take the exam. In this course, approximately 88% of the learners (1892/2161) dropped out; a similarly high dropout rate is frequently reported in the literature (e.g., Jordan, 2014). Therefore, this study explores how the 1892 learners who are considered as dropouts behaved or learned in the selected MOOC. As shown in Fig. 2, more than half of the dropouts are male, and the majority of the dropouts have a degree in higher education (bachelor's, master's or doctoral degree), thus sharing similar patterns with previous studies (e.g., Chen, Sonnert, Sadler, Sasselov, & Fredericks, 2020; Zeng et al., 2020).

2.2. Open-flow network of collective attention

The classical social network is a closed model that fails to account for the high rates of attrition and highly unequal participation patterns of learners. This study built upon our earlier research (Zhang et al., 2019) and adopted the open-flow network of collective attention (Zhang & Wu, 2013) to model learners' behaviours. For a comprehensive review of collective attention, refer to Zhang et al. (2019). This network model uses nodes to represent learning resources and links to represent the learners' sequential visits across these resources. At the collective level, the large body of learners' sequential visits resembles the flux of attention that flows into and out of the learning resources. The flux of such an attention flow forms a network, to which two artificial nodes—'source' and 'sink' —were added to represent the offline space. By adding these two artificial nodes, this network was rebuilt as an open and balanced model, which allows collective attention to flow in and out across online and offline spaces. For individual learning resources, the inflow of attention equals the outflow of attention.

2.2.1. Flow distance and the hierarchical clustering method

In such a collective attention network, flow distance (Guo et al., 2015), which measures the average first-arrival distance between nodes by using the Nth-order Markov transition, was used to calculate the probability that attention would flow in or out of a given learning resource. The flow distance between two nodes is directed (i.e., the flow distance from i to j is not equal to the flow distance

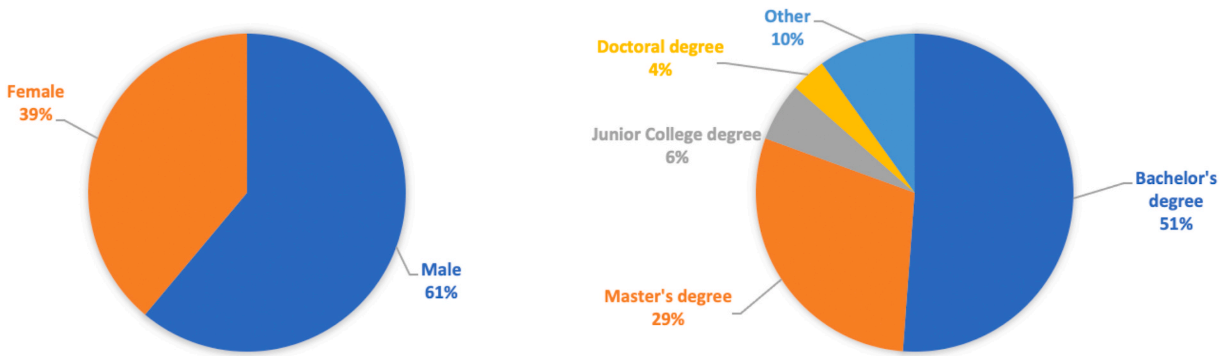

Fig. 2. Distributions of dropouts' gender (left) and educational levels (right).

from j to i). Thus, the symmetric flow distance was used in the hierarchical clustering method (Johnson, 1967) to categorise the learning resources. This method interactively aggregates learning resources that have the shortest distances into a cluster to eventually form one dendrogram, in which all the learning resources are placed in a hierarchical tree structure. This method is applied to understand the patterns of the learning trajectories across different types of resources in an online study.

2.2.2. Allometric scaling relationship

Allometric scaling (power-law) relationships between size and rate are considered the general feature of networks in examining how size supports a particular volume of flow (Banavar, Maritan, & Rinaldo, 1999). A large number of studies have examined allometric scaling in various flow networks, e.g., river basins (Rodríguez-Iturbe & Rinaldo, 2001), food webs (Cohen, Briand, & Newman, 2012), and electrical currents, as well as sewage, prey, nutrient, transportation (Banavar et al., 1999), and vascular networks (West, Brown, & Enquist, 1997). In the collective attention network, I_i is the inflow of the collective attention into learning resource i, which is balanced with the outflow of attention from i. Thus, I_i represents the through flow of collective attention for node *i* (Patten, 1981), that is, the size of a given node *i*. C_i is defined as the sum of $\sum I_i$ on the subtree rooted from *i* (Garlaschelli, Caldarelli, & Pietronero, 2003), which represents the volume of collective attention that can be supported by the size of a given node i. To calculate C_i in a flow network (which is not a tree structure), Patten's method (Patten, 1981, 1982) was adopted. As the collective attention network is a flow network with a steady state, a fixed C_i was calculated with Markov chain techniques and Patten's network analysis. To test the allometric scaling (power-law) relationship, that is, $C_i \propto I_i^{\eta}$, the power-law relationship between C_i and I_i was transformed into a linear regression using the logarithms of C_i and I_i, that is, $\log C_i \propto \eta \log I_i$.

2.2.3. Skeleton of the collective attention network

The Kruskal minimum spanning tree (MST) (Kruskal, 1956) algorithm was employed to generate the skeleton of the network, which includes all the Courseware resources in the collective attention network. In this tree structure, for any given node, another node with the shortest flow distance was added until all the nodes were on the tree that contains the sum of the flow distances, which is minimal. Python and Gephi were employed for the data analysis and visualisation.

3. Results

3.1. Dropouts' registration and learning behaviours

As shown in Fig. 3 (left), approximately 650 of the dropouts registered before the course started (the first unit resources were released on September 10, 2018), and approximately 400 of the dropouts registered after the course ended. The number of participants increased over time, and half of the learners registered before the midterm (indicated with a horizontal green line). This finding, to some extent, reflects that learners chose to learn at their own pace. As shown in Fig. 3 (right), more than 500 learners had already dropped out before the first unit was offered, and approximately 1500 participants left before the last unit was provided.

As a general rule, online behaviours that occurred more than 25.5 min apart were considered as separate sessions (Catledge & Pitkow, 1995). In this study, 30 min was used as a threshold to map out a new session alongside the behavioural sequences. For each session, the session length was applied to measure the number of click behaviours. As shown in Fig. 4 (left), the session lengths varied from one to 62 clicks (as shown on the x-axis), and the distribution of the session lengths exhibited a long tail. This long-tailed

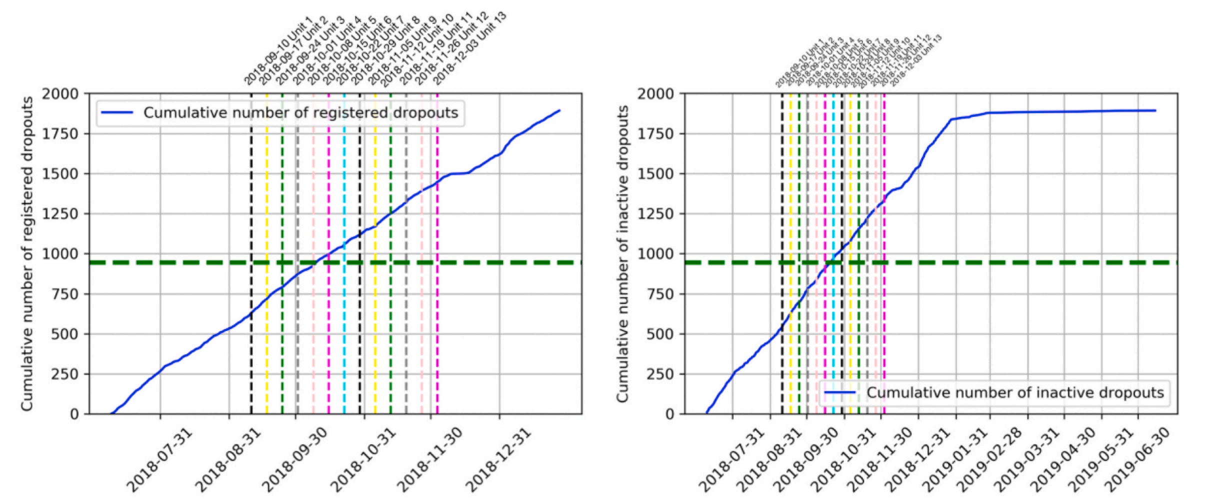

Fig. 3. Pattern of registered dropouts before, during and after the course (left); pattern of inactive dropouts before, during and after the course (right).

distribution includes a large number of sessions that contain only a few behaviours (e.g., 5 clicks), which further confirms that the behavioural trajectories of the dropouts in a single session are short, limited and heterogeneous.

The learning behaviours decrease over the units, which echoes the findings of previous studies (e.g., Yousef, Chatti, Schroeder, Wosnitza, & Jakobs, 2014). As shown in Fig. 4 (middle), the number of participants who accessed units 1–13 also yields a typical long-tailed distribution. While approximately 700 participants accessed Courseware in unit 1, learners dropped out over time; as a result, only 14 learners accessed the unit 13 examination. The behavioural distribution over the unit also has a long tail. unit 1 was accessed 2412 times (mean: 3.34; SD: 3.67), while the last unit attracted only 16 visits.

3.2. More collective attention was allocated to Courseware resources

The visualisation of the collective attention network of the dropouts, which is a directed graph, is presented in Fig. 5 (a simulated random network on the bottom right-hand corner is employed as a null model). Different colours are used to represent different types of learning resources, e.g., Courseware, syllabus, discussion and others. This collective attention network has 70 Courseware resources (video lectures or quizzes, shown as pink nodes), 27 forum resources (shown as blue nodes), and 39 other resources (shown as green nodes). In total, 142 learning resources were connected to each other via the inflow and outflow of the collective attention to produce 1280 links. The average degree of the nodes is 9.01, their diameter is 6, and their average path length is 2.7. Although the network density is 0.064, all the nodes constitute one single, weakly connected component; that is, every node is reachable from every other node disregarding direction. There is no interlinkage that causes the nodes to become several independent subgraphs. Including all the resources from Courseware and the nodes labelled with a #, 125 nodes constitute one strongly connected component in which every node is reachable from other nodes. This outcome further confirms that learning occurred, and it is worth investigating how dropouts learn, considering them as a group from the collective perspective.

Although many resources appeared to belong to the 'other' category (e.g., refund or Sendcloud), minimal attention was allocated to this category (4%). An independent-samples Kruskal-Wallis test revealed statistically significant differences among the attention allocated to Courseware, the discussion forum and other (Kruskal-Wallis $H = 65.473$, $df = 2$, $p = 0.000 < 0.05$). More of the collective attention was allocated to Courseware resources, with a mean rank of 93.62.

3.3. Differentiating between the roles of the about course section and the bulletin resources

The size of the node is proportional to the amount of attention flow into/out of this learning resource. The weight of the link represents the amount of attention flow into/out of occurring between two resources. The two artificial nodes source and sink represent offline space, where the collective attention flows in from the source to the various online learning resources and eventually flows out to the sink (Zhang et al., 2019). The about course node was the most frequently viewed resource (1797 of the dropouts allocated their attention to it 6494 times), and 1016 of the dropouts did not access Courseware after they viewed the about course resource, which is illustrated by the rather strong weight of the links from the source to the about course node and from the about course to the sink node. Only 871 of the dropouts accessed Courseware; these dropouts warrant a detailed examination. In addition to strongly connecting to the about course resource, the two artificial nodes (i.e., source and sink) were also linked to various other learning resources, such as Courseware and the discussions, which suggests that individual dropouts presented heterogeneous learning behaviours. As seen in Fig. 5, Courseware 1.01 in unit 1 also had a role in introducing the course (dropouts allocated attention to this Courseware resource 959 times), and the course bulletin (information page) helped the participants learn about the course requirements. Notably, very few of the dropouts shifted their attention from the course bulletin to Courseware 1.01 or vice versa, but more attention was shifted from the course bulletin to various Courseware resources in the later units in the course. This finding may suggest that some dropouts are likely to learn about course requirements before they advance to learning on Courseware in different units. This suggestion is further confirmed by the result that the info node had the greatest betweenness centrality (3147.25). Although the amount of attention allocated to the course bulletin was less than the amount given to the about course section, the larger betweenness centrality implies that more of the collective attention passed through the course bulletin before or after the participants studied on Courseware.

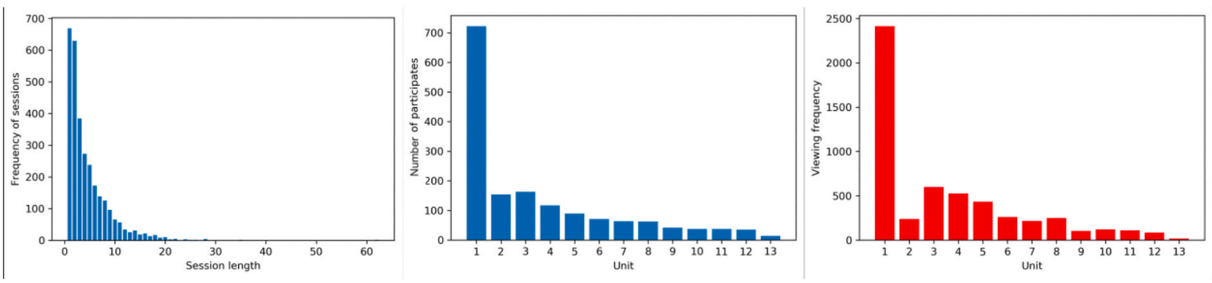

Fig. 4. Frequency distribution of session length (left), number of participants who accessed Courseware across units (middle), and behavioural frequency of learners across units (right).

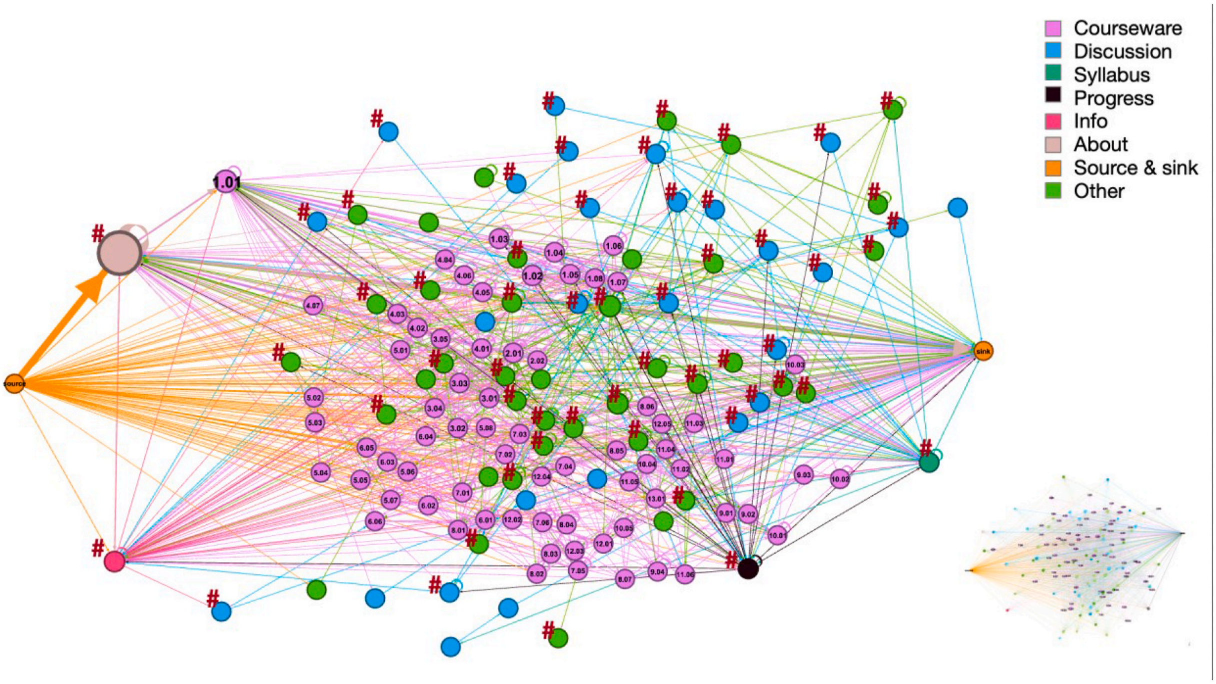

Fig. 5. Open-Flow Network of the Collective Attention (simulated randomised network in the lower right-hand corner).

3.4. *Learning occurred via the forum discussion*

In Fig. 6, the top dendrogram is the result of the clusters of the learning resources modelled by the real learning data, and the bottom dendrogram is modelled by simulating a randomised network (refer to Fig. 5, bottom right) as a control group. Both dendrograms form 5 clusters, among which three clusters in the real network (C1, green branch; C2, red branch; and C4, purple branch) and four clusters in the random network (C1, green branch; C2, red branch; C3, blue branch; and C5, yellow branch) consist of the majority of the resources from Courseware with some resources from the discussion forum and other. This behaviour represents the typical learning behaviour of the dropouts, who learn from videos on Courseware and interact on the discussion forum. One cluster (C5, yellow branch, in the real network and C4, purple branch, in the random network) is filled with all the learning resources, that is, from Courseware and the discussion, about, progress, syllabus, info, and other nodes. This cluster represents the typical behaviours of the dropouts, who sought information about the course requirements and glanced at Courseware and the discussion forum.

In contrast to the randomised dendrogram (bottom, Fig. 6), one cluster (C3, blue branch) in the real data dendrogram mainly consists of the discussion and other nodes, which illustrates that some of the dropouts accessed only the online forum without studying the Courseware content and then went offline. This finding presents a rather different image of dropouts, who used to be viewed as failing to continue, even when micro videos were embedded in Courseware.

3.5. *Dropouts tended to minimise the cost of their collective attention by assigning greater importance to certain resources*

To test whether the allometric scaling pattern observed in the resulting collective attention network was significant, a null model simulated by a random flow network, in which the flows circulate between nodes, was randomly assigned to achieve the same learning path for each learner in one session and ensure that the maximum flux of flow equals that of the original network (Zhang & Wu, 2013).

Using the logarithmic axes to plot C_i and I_i, a linear relationship was identified, which indicates that C_i and I_i have a power-law relationship. As shown in Fig. 7, the red line is a good fit to the power law with slopes of 1.07 for the real network and 0.92 for the random network (refer to Fig. 7). Statistically, the minimum square errors ($R^2 = 0.97$ and $R^2 = 0.88$) indicate a very good fit.

Scaling exponent η indicates the degree of centralisation of the network (Zhang & Wu, 2013). When $\eta > 1$, the network structure is hierarchical (or centralised); that is, some nodes direct the flows of the network. For example, the learning resources with a larger C_i are about course ($C_i = 10{,}751$), unit 1.01 ($C_i = 3205.64$), and bulletin ($C_i = 1631.17$). The majority of the learning resources in unit 1 have a large C_i, including quiz 1.08 ($C_i =$) 507.31. The capacity of these nodes to direct the circulation of the collective is represented by C_i rather than the inflow to these nodes (i.e., A_i). Thus, only qualifying the amount of attention flow fails to illustrate how individual learning resources impact the circulation of the collective attention at the collective level. When $\eta < = 1$, the network structure is decentralised to a large extent, which indicates that each learning resource has similar roles in the network. When the value of η is larger, the flow structure is more hierarchical (centralised) (Wu & Zhang, 2013). The exponent η that is closest to 1 implies that the hierarchy of the collective attention network is low and that this network consists of few levels, which echoes patterns identified in

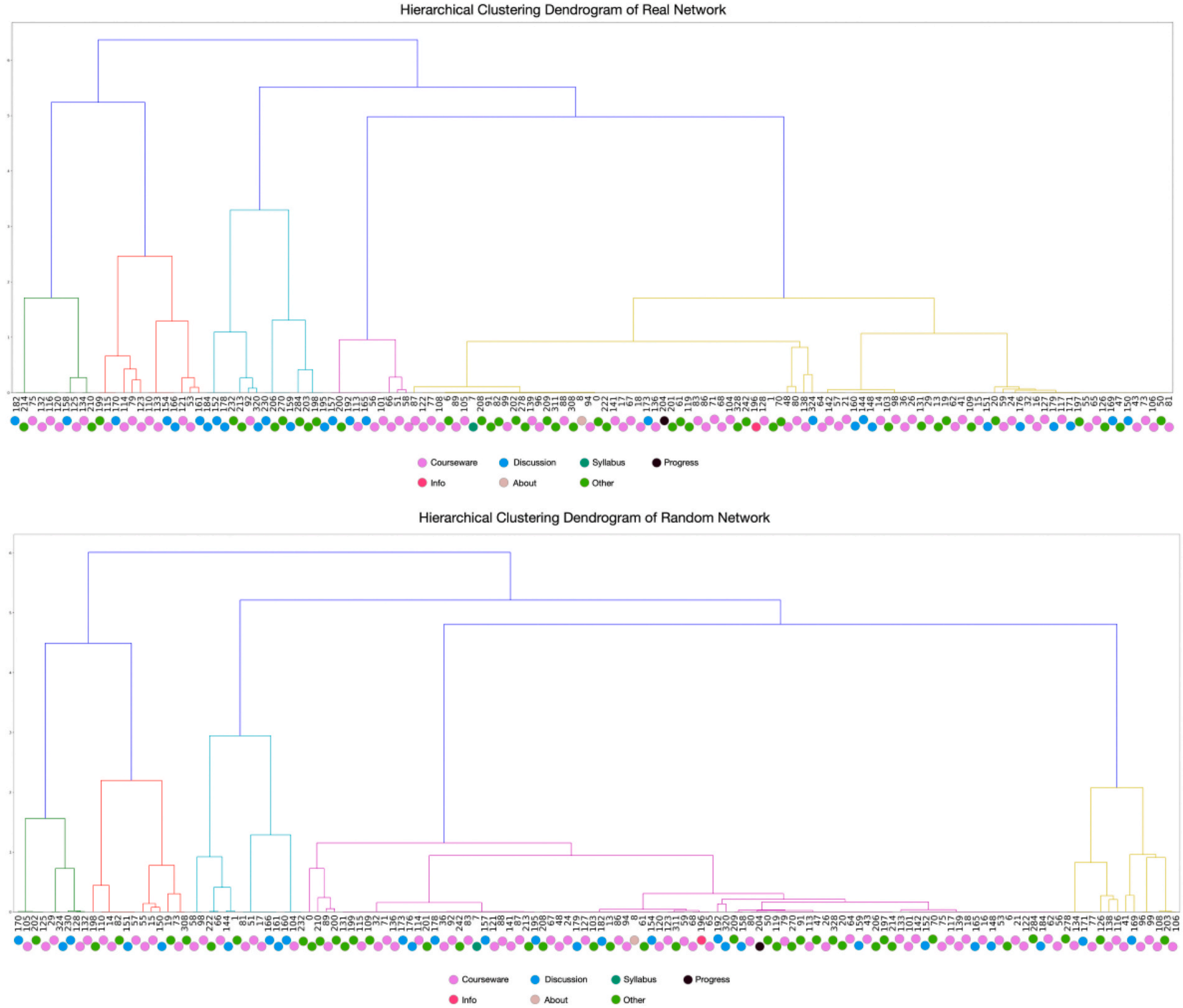

Fig. 6. Hierarchical clustering dendrogram of the real (top) and random (bottom) networks.

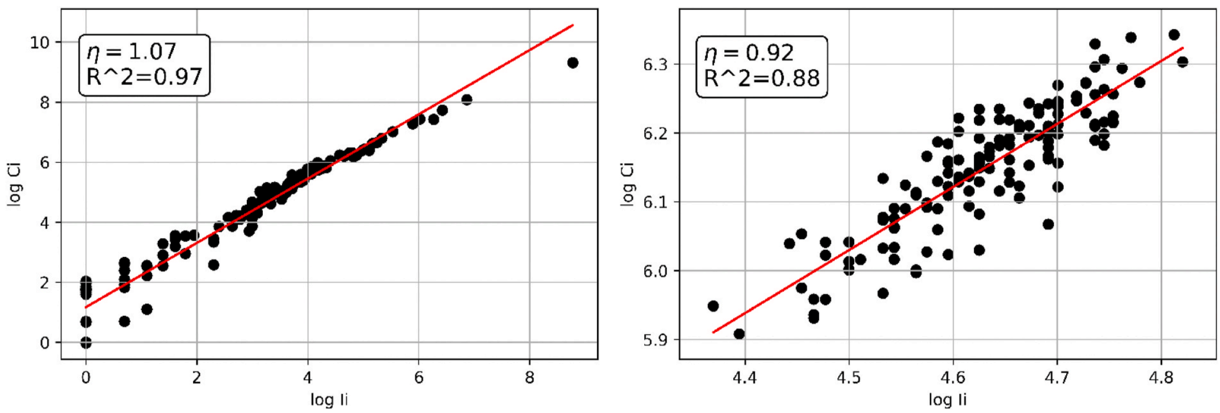

Fig. 7. Allometric scaling pattern in the collective attention network: empirical real network (left); simulated randomised network (right).

other networks (e.g., food webs).

Compared with the null model, $\eta = 1.07$ (>1) indicates that the empirical flow network structure is centralised (Zhang & Wu, 2013); that is, some resources directed the collective attention flow in the network. The dropouts' learning trajectories did not follow a predesigned tree structure of units; instead, the dropouts attended more to some of the learning resources that had a greater role in directing the circulation of the collective attention. As a result of the learners' competing attention, not every resource could be equally attended to, and some of the resources received more of the collective attention.

In this collective attention network, the value of the scaling exponent η also measures the efficiency of the circulation of collective attention (Garlaschelli et al., 2003). Similar to transportation systems (Banavar et al., 1999), this efficiency is an indicator that measures whether the network topology optimises the circulation of collective attention (i.e., the inflow and outflow of the collective attention given to resource *i*). This measurement of efficiency at the system level differs from the evaluation of efficiency for any single learning resource (Garlaschelli et al., 2003). In contrast to the null model, the empirical collective attention network, which is centralised, displays an optimised degree of efficiency, which means that dropouts tended to minimise the cost of their collective attention by assigning a greater importance to certain resources when learning online.

3.6. Quizzes were used as a guidance strategy to direct the collective attention

As most of the collective attention was allocated to Courseware, we further employed the flow distance and MST to analyse the pattern of dropout learning while using Courseware. Having reduced the complexity of the whole attention network to a simpler undirected network graph, a skeleton of the Courseware network was generated by using Kruskal MST (Kruskal, 1956).

As shown in Fig. 8, the MST gives a clear representation of the resulting Courseware network of collective attention. In this network, the nodes represent the Courseware resources (video lectures and quizzes or exams). Learning resources in the same colour belong to the same unit, and the size of the node is proportional to the amount of attention flow into/out of the learning resource. In this tree structure, the sum of the flow distances between nodes is minimal. Zhang et al. (Zhang et al., 2019) discovered that the amount of attention flow and the flow distance were negatively correlated. Thus, the weight of the link is calculated using the converse of the flow distance between the two nodes, which suggests the likely amount of attention flow in (including direct and indirect) and flow out between two learning resources in such a network.

In general, the MST generates similar clusters of units, which closely resemble the predesigned course structure, informed by the subject knowledge of psychology. Specifically, we determined that, apart from units 2 and 3, dropouts did not merely follow the predesigned course tree structure but accessed resources across the different units. The MST represents a new structure that is formed by using the real behavioural data of dropouts (refer to Fig. 8), which echoes the earlier argument that dropouts learn on demand or with different intentions or motivations (Zheng et al., 2015). This tree represents the topological properties of the collective attention network and yields a lower bound on the cost of collective attention. As argued by Zhang et al. (2019), MOOC learning has a high attention cost; thus, the topological structure of the MST contributes to the design of cost-effective learning resources to prevent learners from becoming overloaded.

The visiting pattern of the dropouts presents a long-tailed distribution, which is also reflected in the MST. As shown in Fig. 8, the nodes that belong to unit 1 are the largest, while node 13.01, the exam, is the smallest. This finding illustrates that the amount of the attention flow decreased from the earlier to the later units. However, the largest node, 1.01, which had the most collective attention flow in and out, is not located in the centre of this MST. This node is an isolated 'leaf' that connects to the exam node in the MST, and the remaining learning resources that belong to unit 1 are clustered as leaves on the other side. Instead, unit 13.01 is the distinct centre of the MST, which implies that the cost of collective attention is minimal by giving the exam a central role in making connections with the other learning resources.

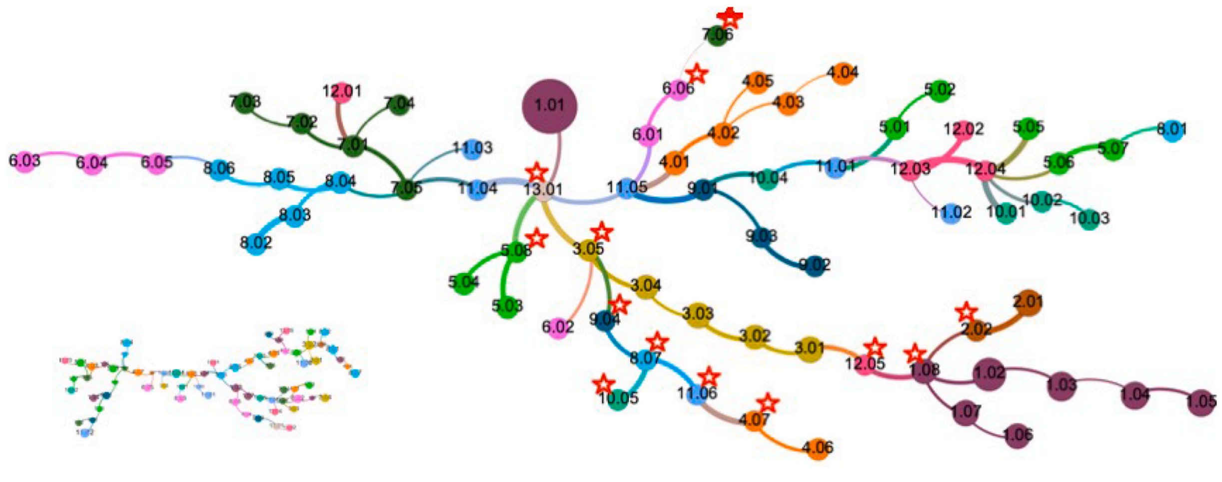

Fig. 8. MST of learners' collective attention flow (randomised MST on bottom left).

We also observe a separation of the learning resources into three large branches of resources and several twigs of varied length. Notably, quizzes are likely to be located at the crossing of the main branches, such as quizzes 1.08, 3.05, 5.08 and 8.07 (marked with red stars in Fig. 8). Several quizzes across different units (e.g., 3, 4, 8, 9, 10, and 11) form a cluster, and further quizzes are likely to act as a bridge between the video resources. For example, quizzes in units 1, 2 and 12 serve as bridges that link units 1, 2 and 3. One possible explanation for this result is that quizzes were not employed by the dropouts to evaluate their studies but that the dropouts used quizzes to guide their learning.

Earlier units are grouped in a line (right corner), which demonstrates similar patterns of following predesigned course structures to learn. The learning units in the midterm—such as units 6, 7 and 8—are dispersed on the other side. Node 8.01, for example, is isolated from the unit 8 cluster, which implies that these learning patterns contrast with those in unit 8, which are situated at the other end of the path. From the perspective of optimisation, learning resources can be restructured beyond subject knowledge to ensure the minimum cost of collective attention and prevent learners from dropping out.

4. Discussions

4.1. *A stereotypical yet different story of how dropouts learn*

Quantifying registration information and behaviours in different units affirmed the stereotype of dropouts as having short, limited and heterogeneous behavioural trajectories while participating in MOOCs. Of the 2161 participants who had complete records in the database, 1892 (88%) dropped out; a similar dropout rate is frequently reported in the literature (e.g., Jordan, 2014). Among the 1892 dropouts, 1016 (54%) accessed only the about course section, and the remaining 871 (46%) accessed Courseware, but these dropouts differ from unsuccessful learners and warrant a detailed examination.

More than half of the dropouts left the course after viewing the about course section, which may suggest that this course was not the course they were seeking or that the dropout's schedule could not accommodated the course. Learners withdraw when they believe their goals will not be achieved or a course does not meet their expectations (Hone & Said, 2016). Previous research has also shown that knowledge of a course is one of the factors that impacts learners' motivation (Deshpande & Chukhlomin, 2017), which has an important role in learners' selection of which courses to take (Hew & Cheung, 2014; Littlejohn et al., 2016).

The 871 dropouts who remained after viewing the about course section allocated more attention to the Courseware resources. It is important for instructors to take seriously their role as content creators to design and develop online resources in Courseware that promote student engagement (Guo, Kim, & Rubin, 2014). Participants are likely to study resources in Courseware and then use the forum discussion as typical MOOC learners, but they also occasionally use the forum discussion only for interaction without watching the videos. This activity presents a rather different image of dropouts, who used to be viewed as failing to continue, even with micro videos embedded in Courseware. Previous research has shown that learners who perform well in courses are better characterised as 'doers' than 'watchers' or 'readers' (Koedinger, Kim, Jia, McLaughlin, & Bier, 2015); our findings suggest that some dropouts can also be doers, which further illuminates the importance of using discussion forums to engage dropouts to weaken the rich-get-richer effect (i.e., Zhang et al., 2016).

4.2. *Distinguishing the roles of introductory learning resources to prevent dropping out behaviour*

An interesting finding is that the course bulletin and Courseware 1.01 had a role in introducing the course that the dropouts began. The course bulletin (info page) helped the participants learn about the course requirements, and the students were likely to check the course requirements every time they logged on to learn, as illustrated by the lack of collective attention between the Courseware 1.01 node and the course bulletin node and by the fact that the info node had the largest betweenness centrality within the network. This finding implies that more of the collective attention circulated through the course bulletin before or after the students learned different Courseware units (especially the later units). This outcome further confirms that these dropouts, to some extent, intended to complete the course, which differs from the assertion that they seek only micro-knowledge in online learning (Ronkowitz, 2013) or that they treat MOOCs as resources rather than designed courses (Zheng et al., 2015).

Compared to previous studies, our research determined that the about course, course bulletin, and introductory video sections seemed to have very different roles in distinguishing among heterogeneous groups of dropouts, who should not be regarded as one, monolithic cohort. Previous research (e.g., Gregori et al., 2018; Sinha, Jermann, Li, & Dillenbourg, 2014) indicated that introductory or prerequisite video lectures or Module 0s are negatively correlated with course completion and hypothesised that this finding might be attributed to the negative impact that overwhelming course information has on learners. Our research further suggests that we should re-examine in depth the roles of the about course, course bulletin and introductory session sections, which have very different roles in predicting who will drop out and instructors should exert extra effort to identify potential dropouts who may need more intensive support to complete the course and pseudo dropouts, whose learning behaviours and characteristics should not be confused with dropouts' real needs and demands. By carefully examining the interaction of the learning behaviours of dropouts with these resources, the number of real dropouts who need learner support can be dramatically reduced, and better design of the course bulletin and introductory video sections might help learners complete the course by meeting the principles of video production proposed by Guo et al. (2014).

Dropouts who frequently check the course requirements before learning should be the targets of an early warning model or a timely intervention strategy. Previous early warning models (e.g., Onah, Sinclair, & Boyatt, 2014) failed to accurately identify potential dropouts worthy of investigation; thus, the use of linear statistical techniques—such as discriminant, logit or probit analyses—has the limitation of considering the interactive effects of independent variables that are introduced by all learners who did not complete courses. MOOC learners are heterogeneous with different educational backgrounds and social, cultural and economic statuses from different parts of the world. Therefore, learning analytics and more advanced machine-learning techniques, such as the extreme learning machine (ELM) (Gregori et al., 2018), are more capable of predicting course completion in MOOCs by considering the high rates of attrition and highly unequal participation patterns.

4.3. Quizzes as a scaffolding strategy to better design MOOCs

The findings of this study also highlight the contrasting role of quizzes in the learning process. As one type of test, quizzes are conventionally utilised to help learners or instructors assess the mastery of knowledge (Marsh, Roediger III, Bjork, & Bjork, 2007). Additionally, tests or quizzes have a role in providing feedback to students based on their strengths and weaknesses and helping instructors improve their teaching (Adesope, Trevisan, & Sundararajan, 2017). However, in this study, learning analytics revealed that quizzes in Courseware tend to guide learners to decide which videos to study rather than to evaluate learners after they have viewed videos. This revelation echoes the finding of McDaniel et al. (McDaniel, Anderson, Derbish, & Morrisette, 2007) that testing can be used to promote as well as evaluate learning.

As the learning resources offered online continue to increase, learners' attention is scarce, and learning becomes more expensive in terms of its attention cost (Zhang et al., 2019). Because MOOC learners have different motivations, intentions, objects and educational backgrounds, tests or quizzes can be an effective strategy for identifying learners' strengths, weaknesses, and knowledge gaps and determining which Courseware resources they wish to access. Previous studies have shown that learners who receive feedback on tests outperform those who do not (Butler & Roediger III, 2008; Mcdaniel & Fisher, 1991), and, of course, some studies did not observe significant differences when feedback was present or absent (Koller, Ng, & Chen, 2013). Nevertheless, minimal attention has been devoted to the use of quizzes and feedback as a strategy or learning design to prevent learner dropout from MOOCs. For example, productive failure is a learning design that encourages learners to persist in challenging tasks. This learning design has been extensively applied in classrooms to activate students' prior knowledge and help students identify knowledge gaps and recognise deep features (Kapur & Bielaczyc, 2012). However, few studies have adapted this learning design in an online setting, which is odd considering that numerous researchers and practitioners have argued that dropout and low completion rates are notorious problems with MOOCs. In MOOC practice, quizzes are employed only to provide right or wrong answers to learners. These multiple-choice tests are easy to evaluate but fail to provide constructive feedback to engage learners or to act as explicit indicators of which resources to access in the future. Given that quizzes have a scaffolding role for dropouts, the design of feedback on quizzes or tests should include direction, guidance or navigation across resources to avoid cognitive load, as previous research has shown that navigational design impacts student motivation to learn in MOOCs (Deshpande & Chukhlomin, 2017).

4.4. Limitations and future work

While this study illustrated interesting findings, certain limitations are noted. The study used only one round of a MOOC, and other offerings of a course might present different behavioural patterns of dropouts; nevertheless, the stories and lessons learned in this study have provided a sound foundation for future work, which could consider using multiple offerings of the same MOOC or comparing different types of MOOCs. Additionally, the exploration of how dropouts learn in a selected MOOC highlights behavioural patterns using only the model of collective attention. Learning intentions and learners' profiles, as argued in the literature (e.g., Al-Adwan, 2020), likely have a major role in constructing similarities in the patterns of learning behaviours, which could be further explored in future work. How to incorporate non-structural information—such as cost, capacity, and time—into the model of collective attention will also be seriously considered in our future work.

5. Conclusion

This study adopted a learning analytics approach using dropouts' heterogeneous learning behaviours to model an open-flow network of collective attention. Taking the systematic and network perspective, this study illustrated a stereotypical but different story about how dropouts learn. This study argued that recognising the distinct roles played by introductory learning resources could prevent dropouts and improve the accuracy of prediction models. Interestingly, assessments embedded in the MOOCs had a scaffolding role in guiding dropouts to learn. Thus, redesigning quizzes or examinations in open and flexible learning environments to construct a minimum cost network of collective attention is vital to making this online space cost-effective for learners at risk. The findings of this study contribute to the field by serving as pragmatic guidance for designing cost-effective learning resources to prevent learners from becoming overloaded. This problem is not trivial in designing an open and flexible learning environment, and it is important to construct a minimum cost network of collective attention that satisfies the prescribed predictors and factors.

Credit author statement

Jingjing Zhang: Conceptualization, Methodology, Writing – original draft; Investigation; Writing- Reviewing and Editing, Project administration, Funding acquisition. Ming Gao.: Data curation, Investigation Writing- Original draft preparation, Methodology, Formal analysis. Jiang Zhang: Methodology.

Statements on open data, ethics and conflict of interest

The clickstream data were provided by XuetangX for research purposes and anonymised. The clickstream data can be provided to readers via a secure server. The clickstream data have not been used for purposes other than originally intended, i.e., to contribute to scholarly research. There is no conflict of interest. The authors have included some of their own work in the analysis according to the criteria outlined in the article.

Acknowledgements

This work was supported by the Chinese National Science Foundation [Project No. 61907004], and the International Joint Research Project of Huiyan Internatinoal College, Faculty of Education,Beijing Normal University. The authors would like to thank XuetangX for providing us clickstream data for this research work.

References

Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *The Internet and Higher Education, 45*, 100724.
Abeer, W., & Miri, B. (2014). Students' preferences and views about learning in a MOOC. *Procedia - Social and Behavioral Sciences, 152*, 318–323.
Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In *Thirty fourth international conference on information systems* (pp. 1–21). Milan.
Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701.
Al-Adwan, A. S. (2020). Investigating the drivers and barriers to MOOCs adoption: The perspective of TAM. *Education and Information Technologies, 25*(6), 5771–5795.
Al-Adwan, A. S., & Khdour, N. (2020). Exploring student readiness to MOOCs in Jordan: A structural equation modelling appraoch. *Journal of Information Technology Education: Research, 19*, 223–242.
Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., & Alalwan, N. (2020). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *Journal of Computing in Higher Education, 32*(2), 429–454.
Ang, K. L. M., Ge, F. L., & Seng, K. P. (2020). Big educational data analytics: Survey, architecture and challenges. *IEEE Access, 8*, 116392–116414.
Banavar, J. R., Maritan, A., & Rinaldo, A. (1999). Size and form in efficient transportation networks. *Nature, 399*(6732), 130–132.
Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. *Computers & Education, 94*, 49–60.
Bonk, C. J., Zhu, M., Kim, M., Xu, S., Sabir, N., & Sari, A. R. (2018). Pushing toward a more personalized mooc: Exploring instructor selected activities, resources, and technologies for mooc design and implementation. *International Review of Research in Open and Distributed Learning, 19*(4), 92–115.
Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment, 8*, 13–25.
Brown, M., Costello, E., Donlon, E., & Giolla-Mhichil, M. N. (2015). A strategic response to MOOCs: How one European university is approaching the challenge. *The International Review of Research in Open and Distributed Learning, 16*(6), 98–115.
Butler, A. C., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.
Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN Systems, 27*(6), 1065–1073.
Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D., & Fredericks, C. (2020). The impact of student misconceptions on student persistence in a MOOC. *Journal of Research in Science Teaching, 57*(6), 879–910.
Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D. D., Fredericks, C., & Malan, D. J. (2020). Going over the cliff: MOOC dropout behavior at chapter transition. *Distance Education, 41*(1), 6–25.
Ciampaglia, G. L., Flammini, A., & Menczer, F. (2015). The production of information in the attention economy. *Scientific Reports, 5*(9452), 1–6.
Cohen, J. E., Briand, F., & Newman, C. M. (2012). *Community food webs: Data and theory* (Vol. 20). Springer Science & Business Media.
Dai, H. M., Teo, T., Rappa, N. A., & Huang, F. (2020). Explaining Chinese university students' continuance learning intention in the MOOC setting: A modified expectation confirmation model perspective. *Computers & Education, 150*, 103850.
Deshpande, A., & Chukhlomin, V. (2017). What makes a good mooc: A field study of factors impacting student motivation to learn. *American Journal of Distance Education, 31*(4), 275–293.
Eom, Y. H., Puliga, M., Smailovic, J., Mozetic, I., & Caldarelli, G. (2015). Twitter-based analysis of the dynamics of collective attention to political parties. *PloS One, 10*(7), 1–17.
Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 51–58). ACM.
Filius, R. M., de Kleijn, R. A. M., Uijl, S. G., Prins, F. J., van Rijen, H. V. M., & Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education, 125*, 86–100.
Garlaschelli, D., Caldarelli, G., & Pietronero, L. (2003). Universal scaling relations in food webs. *Nature, 423*(6936), 165–168.
Garreta-Domingo, M., Sloep, P. B., & Hernández-Leo, D. (2018). Human-centred design to empower "teachers as designers. *British Journal of Educational Technology, 49*(6), 1113–1130.
Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a massive open online course. *American Educational Research Journal, 52*(5), 925–955.
Gregori, E. B., Zhang, J., Galván-Fernández, C., & Fernández-Navarro, F. de A. (2018). Learner support in MOOCs: Identifying variables linked to completion. *Computers & Education, 122*.
Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41–50). New York: ACM.
Guo, L., Lou, X., Shi, P., Wang, J., Huang, X., & Zhang, J. (2015). Flow distances on open flow networks. *Physica A: Statistical Mechanics and its Applications, 437*, 235–248.
Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention–behavior gap. *Distance Education, 38*(3), 353–368.
Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology, 47*(2), 320–341.
Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58.
Hillaire, G., Iniesto, F., & Rienties, B. (2017). Toward emotionally accessible massive open online courses (MOOCs). In *Proceedings of14th AAATE congress 2017*.
Hone, K. S., & Said, G. R. El. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education, 98*, 157–168.

Jiang, S., Williams, A. E., Schenke, K., Warschauer, M., & Dowd, D. O. (2014). Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th international conference on educational data mining (EDM)* (pp. 273–275).

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241–254.

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *International Review of Research in Open and Distance Learning, 15*(1), 133–160.

Kammenhuber, N., Luxenburger, J., Feldmann, A., & Weikum, G. (2006). Web search clickstreams. In *Proceedings of the ACM SIGCOMM internet measurement conference* (pp. 245–250). IMC, 001907.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83.

Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in massive open online courses (MOOCs): The use of learning analytics to reveal student categories. *Journal of Computing in Higher Education, 29*(1), 114–132.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *L@S 2015 - 2nd ACM conference on learning at scale* (pp. 57–66).

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings ofthe 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 60–65).

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 111–120).

Koller, D., Ng, A., & Chen, Z. (2013). *Retention and intention in massive open online courses: In depth*. Retrieved from https://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses-in-depth.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society, 7*(1), 48.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research & Development, 59*(5), 593–618.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education, 29*, 40–48.

Liu, J. G., Yang, Z. H., Li, S. N., & Yu, C. R. (2018). A generative model for the collective attention of the Chinese stock market investors. *Physica A: Statistical Mechanics and its Applications, 512*, 1175–1182.

Li, Z., & Tang, X. (2020). Dynamics of online collective attention as hawkes self-exciting process. *Open Physics, 18*(1), 6–13.

Marsh, E. J., Roediger, H. L., III, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4–5), 494–513.

Mcdaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192–201.

Ministry of Education of the People's Republic of China, M. of E. of the P. R. of C. (2018). Ministry of Education launches 490 national selected online open courses. Retrieved October 20, 2019, from http://en.moe.gov.cn/news/press_releases/201801/t20180119_325124.html.

Olivé, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D. (2020). A supervised learning framework: Using assessment to identify students at risk of dropping out of a MOOC. *Journal of Computing in Higher Education, 32*(1), 9–26.

Onah, D. F., Sinclair, J. E., & Boyatt, R. (2014). Exploring the use of MOOC discussion forums. In *Proceedings of London international conference on education* (pp. 1–4).

Oudeweetering, van de K., & Agirdag, O. (2018). Demographic data of MOOC learners: Can alternative survey deliveries improve current understandings? *Computers & Education, 122*, 169–178.

Patten, B. C. (1981). Environs: The superniches of ecosystems. *American Zoologist, 21*(4), 845–852.

Patten, B. C. (1982). Environs: Relativistic elementary particles for ecology. *The American Naturalist, 119*(2), 179–219.

Poquet, O., Hecking, T., & Chen, B. (2020). The modeling of digital learning networks. In *Companion proceedings 10th international conference on learning analytics & knowledge (LAK20)* (pp. 348–351).

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning, 32*, 202–217.

Ramesh, A., Goldwasser, D., Huang, B., Daum, H., & Getoor, L. (2013). Modeling learner engagement in MOOCs using probabilistic soft logic. In *NIPS workshop on data driven education* (pp. 1–7).

Rasheed, R. A., Kamsin, A., Abdullah, N. A., Zakari, A., & Haruna, K. (2019). A systematic mapping study of the empirical MOOC literature. *IEEE Access, 7*, 124809–124827.

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science, 363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Rodríguez-Iturbe, I., & Rinaldo, A. (2001). *Fractal river basins: Chance and self-organization*. Cambridge University Press.

Romero-Rodriguez, L. M., Ramirez-Montoya, M. S., & Gonzalez, J. R. V. (2019). Gamification in MOOCs: Engagement application test in energy sustainability courses. *IEEE Access, 7*(2), 32093–32101.

Ronkowitz, K. (2013). Rethinking lurkers in the MOOC experience. Retrieved from http://www.serendipity35.net/index.php?/archives/2826-Rethinking-Lurkers-in-the-MOOC-Experience.html.

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., et al. (2014). Social factors that contribute to attrition in MOOCs. In *L@S 2014 - proceedings of the 1st ACM conference on learning at scale* (pp. 197–198).

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM, 57*(4), 58–65.

Shapiro, H. B., Lee, C. H., Wyman Roth, N. E., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education, 110*, 35–50.

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 3–14). EMNLP.

Sunar, A. S., White, S., Abdullah, N. A., & Davis, H. C. (2017). How learners' interactions sustain engagement: A MOOC case study. *IEEE Transactions on Learning Technologies, 10*(4), 475–487.

Sun, N., Rau, P. P. L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior, 38*, 110–117.

Tang, H., Xing, W., & Pei, B. (2018). Exploring the temporal dimension of forum participation in MOOCs. *Distance Education, 39*(3), 353–372.

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology, 46*(3), 472–487.

Tinto, V. (1975). Dropout from higher Education : A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125.

Tschofen, C., & Mackness, J. (2012). Connectivism and dimensions of individual experience. *International Review of Research in Open and Distance Learning, 13*(1), 124–143.

Wang, W., Guo, L., He, L., & Wu, Y. J. (2019). Effects of social-interactive engagement on the dropout ratio in online learning: Insights from MOOC. *Behaviour & Information Technology, 38*(6), 621–636.

Watted, A., & Barak, M. (2018). Motivating factors of MOOC completers: Comparing between university-affiliated students and general participants. *Internet and Higher Education, 37*, 11–20.

Wen, M., Yang, D., & Rosé, C. P. (2014). Linguistic reflections of student engagement in massive open online courses. In *Proceedings of the 8th international conference on weblogs and social media, ICWSM* (pp. 525–534). *2014*.

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science, 276*(5309), 122–126.

Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. In *Proceedings of the national academy of sciences* (Vol. 104, pp. 17599–17601).

Wu, L., & Zhang, J. (2013). The decentralized flow structure of clickstreams on the web. *European Physical Journal B, 86*(6), 1–6.

Yang, D., Sinha, T., Adamson, D., & Rose, C. (2013). "Turn on, tune in, drop out": Anticipating student dropouts in massive open online courses. In *Proceedings of the NIPS workshop on data driven education* (pp. 1–8).
Yang, D., Wen, M., Howley, I., Kraut, R., & Rosé, C. (2015). Exploring the effect of confusion in discussion forums of massive open online courses. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 121–130).
Yang, D., Wen, M., & Rosé, C. P. (2014). Peer influence on student attrition in massively open online courses. In *Proceedings of the 7th international conference on educational data mining, EDM* (Vol. 2014, pp. 405–406).
Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M., & Jakobs, H. (2014). The state of MOOCs from 2008 to 2014: A critical analysis and future visions. In S. Zvacek, M. T. Restivo, J. Uhomoibhi, & M. Helfert (Eds.), *Proceedings of the international conference on computer supported education* (pp. 305–327). Cham, Switzerland: Springer.
Zeng, S., Zhang, J., Gao, M., Xu, K. M., & Zhang, J. (2020). Using learning analytics to understand collective attention in language MOOCs. *Computer Assisted Language Learning*.
Zhang, J., Lou, X., Zhang, H., & Zhang, J. (2019). Modeling collective attention in online and flexible learning environments. *Distance Education, 40*(3), 303–308.
Zhang, J., Skryabin, M., & Song, X. (2016). Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis (SIENA). *Distance Education, 37*(3), 270–286.
Zhang, J., & Wu, L. (2013). Allometry and dissipation of ecological flow networks. *PloS One, 8*(9), 1–8.
Zhang, J., & Yang, Y. (2019). The power law in online learning: New index based on open and balanced system. *The Journal of Distance Education*, (4), 96–105.
Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Understanding student motivation, behaviors, and perceptions in MOOCs. In *CSCW 2015 - proceedings of the 2015 ACM international conference on computer-supported cooperative work and social computing* (pp. 1882–1895).