

# Participation and Interaction Patterns of MOOC Learners with Different Learning Achievements: A Collective Attention Network Perspective

Ming Gao <sup>1</sup>, Jingjing Zhang <sup>1,2,3</sup>

<sup>1</sup> Research Centre of Distance Education, Beijing Normal University, Beijing, China

<sup>2</sup> Big Data Centre for Technology-mediated Education, Beijing Normal University, Beijing, China

<sup>3</sup> Faculty of Education, Beijing Normal University, Beijing, China

## Abstract

MOOCs offer a wide range of students an open and cost-free learning environment as a kind of online education. Understanding how students allocate their attention among a growing number of well-designed educational resources is important to making sense of this type of online and flexible learning. This study selected the ‘Introduction to Psychology’ offered on XuetangX as an example, and adopted an open-flow network approach to investigate how different learning achievement groups allocate their attention at a collective level. The results showed that the patterns of collective attention differed among different learning achievement groups. Learners who received excellent and good scores were more likely to attend to the course syllabus and learning progress, whilst learners who failed or did not participate in the exam were more likely to wander around. Moreover, the cost of collective attention while attending to different resources follows the pre-designed course structure, which illustrates that learners are largely manipulated by the pre-designed tree structure, which packages knowledge into separate categories. As 21<sup>st</sup> century skills demand learners to make the connection between knowledge, how to take full advantage of the open and flexible nature of the online course to facilitate more personalized and complex interaction between learners and content deserves severe attention.

## Keywords

MOOCs, collective attention, open-flow network, network science, learning analytics

## 1. Introduction

Massive open online courses, known as MOOCs, are becoming increasingly popular among students looking for open educational opportunities online. When compared to traditional classrooms with a finite number of students, MOOCs offer a more open, flexible, and affordable way for a far larger number of individuals to learn [1]. Learners of MOOCs are not required to pay any fees if they do not wish to pursue the option of receiving a course certificate. As a result, MOOCs appear to be no cost, which is in keeping with the open and free philosophy that has long been promoted. Therefore, it is commonly argued that when students have such unfettered access to learning resources of a high quality, they will make effective use of those resources. However, when students are presented with a vast quantity of resources, the true cost is their scarce and limited attention [2,3]. In a similar vein, although more and more learning resources are being made available in MOOCs, students can only concentrate their attention on a certain number of resources at a time. When it comes to individuals, their attention can be distributed and channeled toward a variety of resources. Individual attention is dispersed, but it is concentrated and formed into collective attention [4], which

---

Proceedings of the NetSciLA22 workshop, March 22, 2022

EMAIL: mgao519@126.com (A.1); jingjing.zhang@bnu.edu.cn (A.2)

ORCID: 0000-0003-0129-703X (A.1); 0000-0002-0584-534X (A.2)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

gives us an opportunity to understand how students allocate their attention to participate and interact in the massive information space on a collective level.

From the perspective of cognition [5], educational psychologists have spent years trying to understand how learners concentrate on or transfer their attention to the instructional content. Unlike traditional classroom learning, learners need a good self-regulatory ability to select learning resources, allocate learning time, and set a learning pace in the online learning environment where learning resources are excessive [6]. In this case, learners' attention has been a more macroscopic and external form of resource [7] and their academic attainments are related to the effective accumulation, circulation, and dissipation of collective attention flow [2]. "Attention economy" was introduced by Goldhaber, who defined attention as a scarce resource. He stated that attention may also be traded and flowed in different places in the online environment where we increasingly dwell, just as traditional resources can be traded and flowed to make profits in economics [8]. In this view, human attention is given the attribute of flow creatively. To put it another way, attention moves continuously between the various resources available on the web, both in and out. As a result, there is a cost, which could be very costly when people's attention flows in and out of resources, since there is only so much attention to go around. In this study, we take four rounds of one course on XuetangX as an example and build upon earlier studies by using the network model of collective attention to investigate the participation and interaction patterns of learners with different achievements at the collective level. Clickstream data from four rounds of 'Introduction to Psychology' (Autumn 2017, Spring 2018, Autumn 2018, Spring 2019) were used in the present study to address the following two questions:

1. What do the allocation patterns of collective attention look like among four learning achievement groups?
2. What is the cost of the collective attention flow in and out of courseware resources among different groups?

## 2. Methods

### 2.1. Context

Introduction to Psychology on XuetangX was selected as a case to analyze the patterns of participation and interaction. As one of the top-quality open courses [9], it has launched several rounds since 2015. Referring to the traditional instructional regime, it also set two open accessing periods - spring and autumn semester - for the public each year. The course design adopted a traditional linear trajectory which consists of several units (see Figure 1.a). Each unit provides several video lectures, followed by an assignment for learners to self-evaluate. 70 videos, assignments or examination from 13 units form the main learning resources of the course (Figure 1.b shows the detailed number of the learning resources in each unit). Moreover, the course also provides other types of modules or resources, such as a course forum, syllabus, progress, etc. [10]. The overall course design almost remained in the different rounds.

While learners participate in the course, their learning behavioral data will be automatically collected and stored in the database. In our previous studies, we found that there was a tendency for learners to access this course on mobile devices. Unfortunately, behavioral data created via mobile devices was not stored in a complete form. As a result, we only left the available data from participants who accessed the course via the web. Therefore, there is a sharp decline in the amount of data that can be analyzed in a single round of the course. Fortunately, there is a possible way to combine data from multiple rounds of the course to increase the amount of available data and analyze it as the overall course design has remained the same. In this study, four rounds (Autumn 2017 to Spring 2019) of data were selected.

In these most recent four rounds of the course, a total of 21,182 learners accessed the course via the web and left 155, 191 behavioral records. When looking up the registration information and examination records, only 9,438 of these learners had complete records in the database. After cleaning the data set, 9,438 users who accessed the course via the web had complete records in the database, and at least 1 second of log data (105,160 logs) was included in the data analysis. According to the score they achieved in the final examination (the maximum is 100), we grouped these 9,438 learners into four cohorts. 79 participants whose scores were greater than or equal to 80 were labelled as

excellent; 85 participants whose scores were between 60 and 80 were labelled as good; 1139 participants were labelled as failed as their final scores were lower than 60; and the remaining 8135 participants were labelled as absent because they did not take the exam. Referring to their reported demographic profiles, 936 learners were female, 1,232 were male, and 7270 did not declare their gender. About 80% of those who told about their education had at least a bachelor's degree.

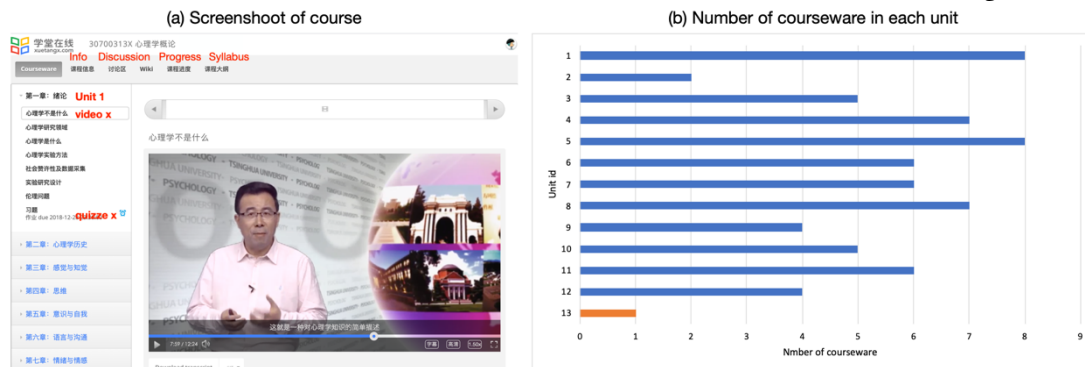


Figure 1: Description of the Course

## 2.2. Modelling an open-flow network of collective attention

Based on the learners' clickstream data and our earlier research [2, 10, 11, 12], open-flow networks of collective attention were created to model different cohorts' behaviors. Learning sessions were used to establish a collective attention network due to the fact that individual learners do not acquire all of the learning units at the same predetermined period. As a rule of thumb, online activities that happened more than 25.5 minutes apart were treated as separate sessions [13]. In this study, a new session was defined as a sequence that had passed 30 minutes.

Figure 2 shows the network model of collective attention and its flow between online and offline space. The nodes in this network model represent learning resources, and the links indicate the learners' sequential visits to these resources. At the collective level, the large body of consecutive visits by learners reflects the flux of attention that flows into and out of the learning resources. The network was formed by the flux of such an attention flow. The offline space was represented by the two added artificial nodes—'source' and 'sink'. The route from the "source" to the other nodes depicts the point at which people get online; when people go offline from one resource, this is marked by a link from the nodes to "sink" (i.e., 'sink' marks the end of a learning session). 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 (more details are given in [2]). For individual learning resources, the inflow of attention equals the outflow of attention. In this study, the total amount of attention flowing into each learning resource was taken into consideration.

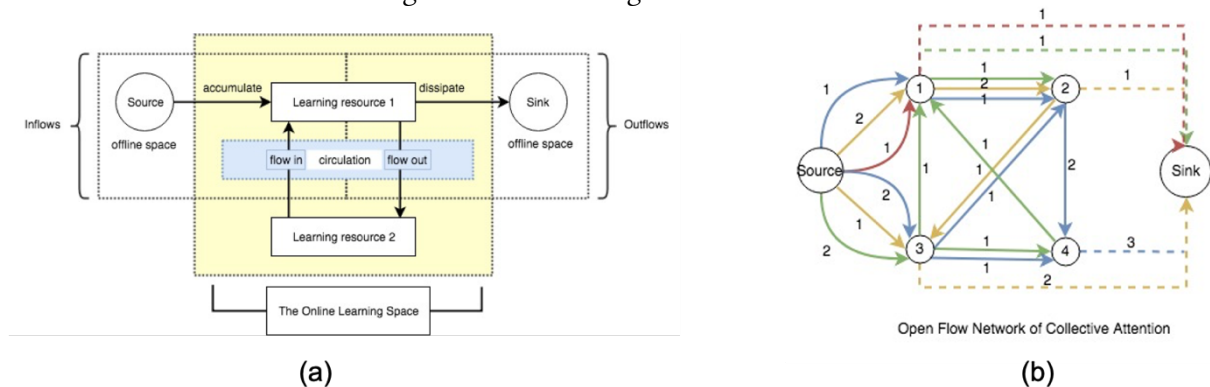


Figure 2: The accumulation, circulation, and dissipation of attention flow (from [2], p. 287)

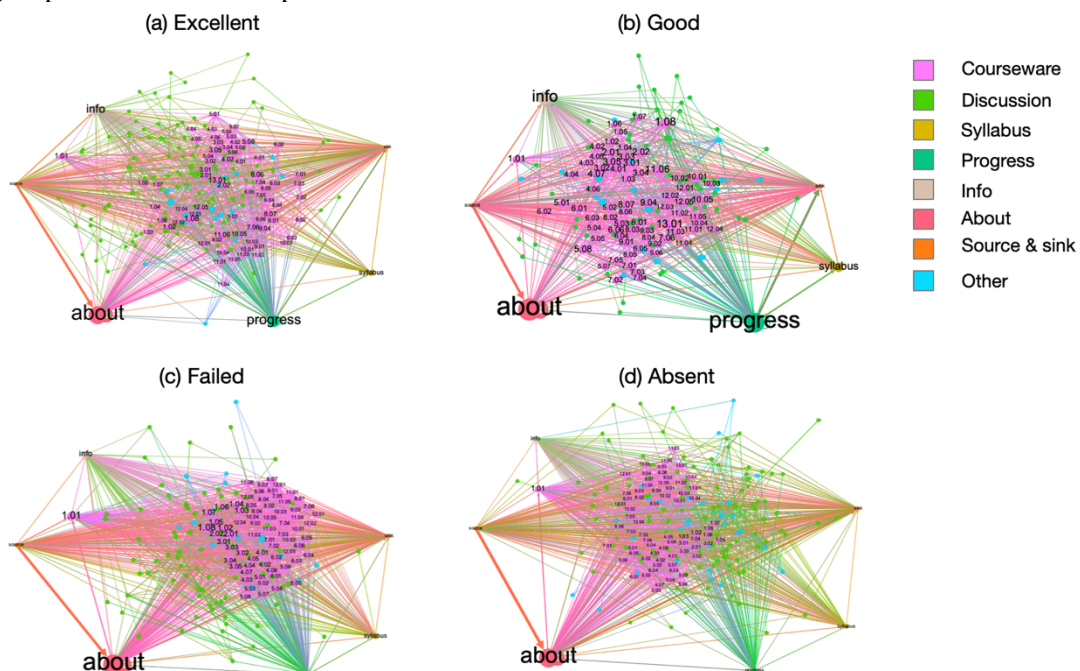
Flow distance, an innovative metric proposed by Guo [14], was used to measure the tendency of attention flow between learning resources. It was defined as the average first-arrival distance

estimated by an N-order Markov transition between nodes. The flow distance from node 1 to node 2 is equal to the sum of the probabilities of learners from node 1 visiting node 2 for the first time after k steps, where k spans from 1 to infinity, according to the infinite order calculation of Markov transfer. The idea depicts the average number of steps necessary for attention flow to circulate from one learning resource to another because flow distance estimates the average first-arrival distance between two nodes of a learning resource. A small flow distance from node A to node B indicates a greater tendency of attention flow direction from node A to node B. The value of flow distance can also be used to measure the cost of collective attention. For example, if the value of the flow distance between the two contents is large, it means only a small part of learners' attention will flow according to the current order or sequence. This indicates a high cost of collective attention flow between these two contents. Namely, learners usually do not allocate their attention to the next piece of content after completing the previous one. On the contrary, if the value of the flow distance is small, it shows that most of the learners' attention flows follow the current order and indicates that the pre-designed learning sequence has a great effect on guiding the flow of collective attention.

### 3. Results

#### 3.1. What do the allocation patterns of collective attention look like among four learning achievement groups?

Based on the learners' clickstream data, four open-flow networks of collective attention were built and are visualized in Figure 3. The different colors are used to represent different types of learning resources, e.g., courseware, syllabus, discussion, and others. All collective attention networks have the same number of courseware resources (video lectures or quizzes, the total number is 70), course bulletin (Info, the number is 1), introduction of the course (about resource, the number is 1), syllabus resource (the number is 1) and learning progress resource (the number is 1). There are some differences in the number of forum resources (discussion) and other resources (other) between the four groups. The node size represents the amount of flow into and out of the collective attention.



**Figure 3 Open-flow networks of collective attention of four groups**

The metrics of the four collective attention networks are shown in Table 1. There are more nodes in the networks of groups of learners who performed excellent or did not participate in the final exam. Moreover, the network densities of these two groups were smaller than the other two groups. That is, these two groups' learners are more likely to attend to more learning resources than the other two

groups, who were used to following pre-designed instruction and whose attention flowed to more learning resources except the course pre-designed normal learning.

For the average degree, a one-way analysis of variance (ANOVA) test revealed a significant discrepancy between the four groups ( $F(3, 720) = 3.697, p = 0.012 < 0.05$ ). The further Turkey multiple comparison test (Tukey's HSD) showed that the failed group's average degree is significantly larger than the excellent group's (means differ = 7.624,  $p = 0.025 < 0.05$ ,  $CI = [0.664, 14.585]$ ), while there were no significant differences between other groups. It indicated that failing learners' attention was more likely to circulate more learning resources than excellent learners'. However, the diameter of the excellent group was larger than the other groups. It may indicate that learners in the excellent group tend to form longer learning paths than those in other groups.

It is interesting to find that the group of learners who did not pass the exam not only has the largest value of node, edge, and network density, but also has a smaller average path length. It means that learners in this group frequently attend to all kinds of resources. One possible explanation is that they try to build the connection between resources or find the answer in any possible way.

**Table 1**  
Attributions of four open-flow networks of collective attention

Group	Num. of nodes	Num. of edges	Average degree	Diameter	Average path length	Network density
Excellent	202	1627	8.054	14	3.320	0.040
Good	150	1229	8.193	6	2.675	0.055
Failed	165	1958	11.867	8	2.370	0.072
Absent	207	2200	10.628	6	2.473	0.052

The amount of collective attention flowed into each learning resource was also calculated and differences between the groups of learners who performed extremely well or good in the exam and the groups of learners who did not pass or participate in the exam were found. The largest amount of attention flows into the course bulletin (Info), introduction of the course (About), syllabus resource (Syllabus), learning progress resource (Progress), and several courseware resources (i.e., 1.08, 13.01, 1.01, 3.05) for the groups of learners who performed extremely well or good in the exam. For the groups of learners who did not pass or participate in the exam, the top 10 learning resources with the largest attention flow are the course bulletin (Info), introduction of the course (About) and several courseware resources (i.e., 1.01, 1.02, 1.03, 1.04, 1.05, 2.01). It could imply that the two former groups, excellent and good, are more concerned with the instructional plan and their learning progress than with directing their attention flow in the course learning.

### 3.2. What is the cost of the collective attention flow in and out of courseware resources among different groups?

In MOOCs, especially in xMOOCs, the main course contents (i.e., courseware provided by the course instructors, such as lectures, quizzes, and examinations) are usually organized as a tree structure, with chapters and units. This categorized structure is an important piece of guidance designed by instructional designers and subject experts. It aims to help learners follow a logical path to learn knowledge as time goes on. However, in an online setting, while accommodating self-regulated learning in a more open and flexible manner, whether learners still strictly follow the illustrated course structure, or whether they should follow the course design, as this group is not homogeneous, are questions that need to be answered. To look into these issues, we aligned the learning resources with the unit structure and used the flow distance as a way to figure out how much attention the course structure took from everyone.

Based on the tree structure of learning resources, learners' ideal learning path can be simplified as follows: (1) learners enter the online course learning space; (2) learners complete the content of all the units in the given order of the course; (3) learners leave the online course learning space. That is,

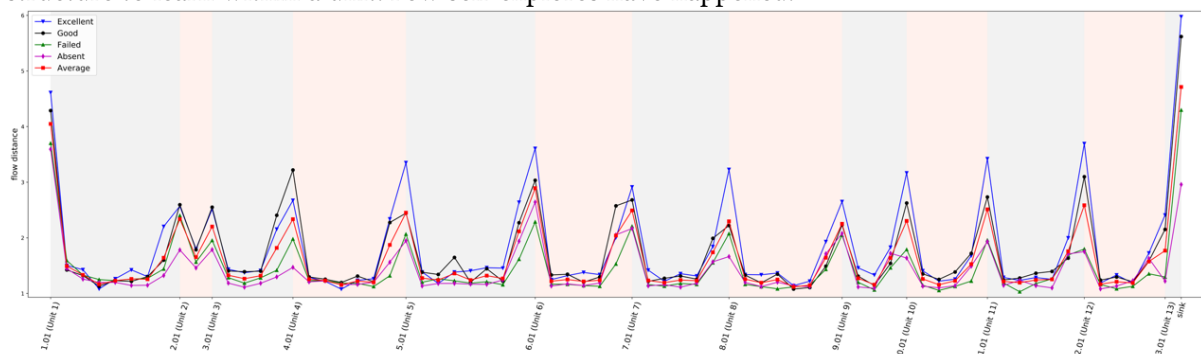
the learning process can be described as the following order: source -> 1.01 -> 1.02 -> ... -> 2.01 -> ... -> 3.01 -> ... -> 13.01 -> sink. The flow distance between two items shows the tendency of collective attention to flow. If the flow distance between the two contents is long, it means only a small part of the learners' attention will flow according to the current linear order. This indicates a high cost of collective attention flow between these two contents. Namely, learners usually do not allocate their attention to the next piece of content after completing the previous one. On the contrary, if the flow distance is short, it shows that most of the learners' attention flows follow the linear order and indicates that the pre-designed learning sequence has a great effect on guiding the flow of collective attention.

Figure 4 shows the results of the cost of collective attention for different groups and their average level. The x-axis shows the sequential learning order in which the first content/node is the first lecture of Unit 1 (i.e., 1.01) and the last content/node is sink, which indicates the offline space. Besides, in the background, two colors (red and gray) are interleaved to form separate units. The y-axis shows the flow distance between two resources. For example, the value of the first node is the flow distance of collective attention from source to 1.01, and the value of the last node is the flow distance of collective attention from 13.01 to sink. Note that if no collective attention flows between two nodes, the flow distance is null. We give it the maximum flow distance to represent the flowing cost of collective attention between these two nodes in order to make it quantifiable and visualized.

As shown in Figure 4, it can be found that four groups present a similar pattern of collective attention cost. First, the flow distances between the last resource of the previous unit and the first resource of the latter unit (i.e., the node located at the intersection of two different color backgrounds) are usually larger, while the flow distances between resources within a unit are smaller. It shows that the flowing cost of collective attention within a unit is small, whilst the flowing cost of collective attention between different units is higher. It indicates that the linear tree structure of the contents manipulated the learners' learning trajectories to some extent. However, there were certain differences in the cohesion of content between the different units, because learners cost more collective attention when they followed the pre-designed linear learning sequence between units.

Second, the flow distances between two adjacent pieces of content (i.e., lectures or quizzes) within a unit often decrease first and then rise. It indicated that the contents within a unit were more related and that it was in line with the learners' cognitive development. Therefore, they usually learned this content by following the pre-designed linear structure.

Third, we also found that within some units, such as unit 9 and unit 12, the flow distances were more fluctuant. It may indicate that learners are not always following the pre-designed linear structure to learn within a unit. Few self-explodes have happened.



**Figure 4** Flowing cost of collective attention of different groups when they followed the pre-designed linear learning sequence

## 4. Discussions and Conclusion

This study attempted to improve our understanding of online learning in an open and flexible learning environment from a network science perspective. To learn about the participation and interaction patterns, we divided learners into four different groups according to their performance in the final examination and created four open-flow networks of collective attention. To examine the

differences in allocation and cost patterns of attention, we compare the network structural properties, the amount of collective attention, and the flow distance of different types of learning resources.

In the top 10 learning resources with the highest traffic of collective attention, we found that About, Info, and several courseware resources are shown in all groups. However, the syllabus and progress resources are only shown in the top 10 learning resources in the groups of learners who perform extremely well and good in the exam. These two types of learning resources (i.e., syllabus and progress) are more related to the progress of an instructional plan and their learning. It may indicate that learners' final performance depends on their attention to the progress of the course and timely learning. This finding is consistent with previous literature [15,16], which verifies that it is important for learners to study regularly and timely in online courses.

Although MOOCs provide students with an open and flexible online learning environment, their attention has become a scarce resource, and they have decided what we will achieve in the face of abundant learning resources [2]. In this study, we also analyzed the cost of collective attention of different groups, especially the attention that was allocated to the instructors' provided courseware resources. We measured the flowing tendency of collective attention between the courseware resources and found that it costs less when collective attention flows within a unit, but more when collective attention flows between different units. Researchers argue that the online learning space gives learners enough freedom so that learners can be free to jump between different learning contents and they will more actively engage in non-linear navigation [17]. However, the flowing cost of collective attention between and within units revealed that learners still prefer to learn following a pre-designed course structure, which is in line with what the previous study argued [18,19]. Therefore, how to take full advantage of the open and flexible nature of online courses to facilitate more personalized and complex learning interactions to prepare students to make the connection between knowledge may be a key issue in online course design.

## References

- [1] B.D. Voss, "Massive Open Online Courses (MOOCs): A Primer for University and College Board Members," 2013. AGB Association of Governing Boards of Universities and Colleges. Review. [https://agb.org/wp-content/uploads/2019/01/report\\_2013\\_MOOCs.pdf](https://agb.org/wp-content/uploads/2019/01/report_2013_MOOCs.pdf)
- [2] J. Zhang et al., "Modeling collective attention in online and flexible learning environments," *Distance Educ.* vol. 40, no. 2, pp. 278–301, 2019.
- [3] H. Simon, "Designing organizations for an information-rich world," *Comput. Commun. Public Interes.*, vol. 72, pp. 38–72, 1971.
- [4] F. Wu and B. A. Huberman, "Novelty and Collective Attention," *Proc. Natl. Acad. Sci.*, vol. 104, no. 45, pp. 17599–17601, 2007.
- [5] D. Kahneman, *Attention and effort*. Englewood Cliffs: PRENTICE-HALL, INC., 1973.
- [6] C. Milligan and A. Littlejohn, "Supporting Professional Learning in a Massive Open Online Course," *Int. Rev. Res. Open Distrib. Learn.*, vol. 15, no. 5, pp. 197–213, 2014.
- [7] M. B. Crawford, *The world beyond your head: On becoming an individual in an age of distraction*. Farrar, Straus and Giroux, 2015.
- [8] M. H. Goldhaber, "The attention economy and the net," *First Monday*, vol. 2, no. 4, 1997.
- [9] Ministry of Education of the People's Republic of China, "Ministry of Education launches 490 national selected online open courses," 2018. [http://en.moe.gov.cn/news/press\\_releases/201801/t20180119\\_325124.html](http://en.moe.gov.cn/news/press_releases/201801/t20180119_325124.html)
- [10] J. Zhang, M. Gao, and J. Zhang, "The learning behaviours of dropouts in MOOCs: A collective attention network," *Comput. Educ.*, vol. 167, p. 104189, 2021.
- [11] S. Zeng, J. Zhang, M. Gao, K. M. Xu, and J. Zhang, "Using learning analytics to understand collective attention in language MOOCs," *Comput. Assist. Lang. Learn.*, 2020. <https://doi.org/10.1080/09588221.2020.1825094>
- [12] J. Zhang, Y. Huang, and M. Gao, "Video Features, Engagement, and Patterns of Collective Attention Allocation: An Open Flow Network Perspective," *J. Learn. Anal.*, vol. 9, no. 1, pp. 32–52, 2022.

- [13] L. D. Catledge and J. E. Pitkow, "Characterizing Browsing Strategies in the World-Wide Web," in *The Third International WWW Conference*, 1995, pp. 1–9.
- [14] L. Guo, X. Lou, P. Shi, J. Wang, X. Huang, and J. Zhang, "Flow distances on open flow networks," *Phys. A Stat. Mech. its Appl.*, vol. 437, pp. 235–248, 2015.
- [15] J. W. You, "Examining the Effect of Academic Procrastination on Achievement Using LMS Data in E-Learning," *Educ. Technol. Soc.*, vol. 18, no. 3, pp. 64–74, 2015.
- [16] C. J. Asarta and J. R. Schmidt, "Access Patterns of Online Materials in a Blended Course," *Decis. Sci. J. Innov. Educ.*, vol. 11, no. 1, pp. 107–123, 2013.
- [17] C. H. M. Lee, F. Sudweeks, and Y. W. Cheng, "A longitudinal study on the effect of hypermedia on learning dimensions, culture and teaching evaluations," in *Proceeding Cultural Attitudes Towards Technology and Communication*, 2012, pp. 146–162.
- [18] N. Ford and S. Y. Chen, "Matching/mismatching revisited: An empirical study of learning and teaching styles," *Br. J. Educ. Technol.*, vol. 32, no. 1, pp. 5–22, 2001.
- [19] C. H. M. Lee, F. Sudweeks, Y. W. Cheng, and F. E. Tang, "The Role of Unit Evaluation, Learning and Culture Dimensions Related To Student Cognitive Style in Hypermedia Learning," in *Proceedings Cultural Attitudes Towards Communication and Technology*, 2010, pp. 400–419.