Gauging MOOC Learners' Adherence to the Designed Learning Path

Dan Davis[*], Guanliang Chen[†], Claudia Hauff and Geert-Jan Houben
Delft University of Technology
Delft, the Netherlands
{d.j.davis, guanliang.chen, c.hauff, g.j.p.m.houben}@tudelft.nl

ABSTRACT

Massive Open Online Course (MOOC) platform designs, such as those of edX and Coursera, afford linear learning sequences by building scaffolded knowledge from activity to activity and from week to week. We consider those sequences to be the courses' *designed* learning paths. But do learners actually adhere to these designed paths, or do they forge their own ways through the MOOCs? What are the implications of either following or not following the designed paths? Existing research has greatly emphasized, and succeeded in, automatically predicting MOOC learner success and learner dropout based on behavior patterns derived from MOOC learners' data traces. However, those predictions do not directly translate into practicable information for course designers & instructors aiming to improve engagement and retention — the two major issues plaguing today's MOOCs. In this work, we present a three-pronged approach to exploring MOOC data for novel learning path insights, thus enabling course instructors & designers to adapt a course's design based on empirical evidence.

Keywords

MOOCs, learning path analysis, visualization

1. INTRODUCTION

MOOCs can deliver a world-class education on virtually any academic or professional development topic to any person with access to the Internet. Millions of people around the globe have signed up to courses offered on platforms such as edX, Coursera, FutureLearn and Udacity. At the same time though, only a very small percentage of these learners actually complete a MOOC successfully [15], an issue that continues to plague massive open online learning. Keeping MOOC learners engaged and improving the dismal retention rates are major concerns to instructional designers and MOOC instructors alike. Considerable research efforts have been dedicated to the automatic prediction of learners' (imminent) dropout in MOOCs, e.g. [9, 12, 17, 24], under the assumption that once learners under the threat of attrition are identified, an automated intervention can be staged to (re)engage those learners with the course material. While the accuracy of these usually machine-learning-based predictors is high, their *explanatory power* is often low. Model features that have the strongest impact on prediction purely based on statistical grounds may not provide course designers & instructors with enough information to adapt the design or content of a MOOC in response.

In this work we aim to provide a more holistic view of learners' progression through a MOOC in order to enable more practicable insights to instructors and designers. Our approach to educational data mining as presented here is a very literal realization of Graesser's vision for the field by illustrating and "*look[ing] at unique learning trajectories of individuals*" [21]. We make use of the concept of *learning paths* (a learner's route through course activities) and investigate how the learning paths of successful and unsuccessful MOOC learners differ.

The design of MOOCs on the edX platform[1] implies a *linear* trajectory through the learning material. Most courses are broken up into weeks (Week 1, Week 2, etc.) and released one week at a time. Within these weeks, the standard instructional approach is to first provide a brief introduction to the week's material, followed by the weekly video lectures (the main source of content delivery), then the assessments that evaluate learners' knowledge of the preceding video lectures, and, finally, courses may offer bonus material. This cycle is repeated each course week (and sometimes multiple cycles comprise a single week). But do learners *actually* adhere to this cycle, and thus the *designed* learning path? Does it matter if they do not? These are the central issues that we focus on in this paper. While the concept of *executed* learning paths (i.e., the paths students actually take through a course) has received substantial attention in the e-learning and intelligent tutoring communities [13, 19], in the MOOC setting this concept has so far garnered little attention. First empirical evidence that learners do not always follow the designed sequence through a MOOC has been observed in [8], however, to our knowledge no in-depth investigation of this phenomenon in the MOOC context exists as of yet. We aim to close this knowledge gap and investigate the following

[*]The author's research is supported by the *Leiden-Delft-Erasmus Centre for Education and Learning.*

[†]The author's research is supported by the *Extension School* of the Delft University of Technology.

[1]Our empirical work is based on edX MOOCs, but the same principles apply to other major MOOC platforms.

research question:

*To what extent do learners **adhere** to a MOOC's designed learning path?*

We develop three approaches to characterize learning paths, thus providing three different views on a MOOC's *designed learning path* (created by the course instructor or designer) and the *executed paths* (created by the learners of the MOOC). We apply our approaches on the log traces of more than 113,000 learners who participated in one of four edX-based MOOCs in the domains of computer science, political debates and business ethics.

We show that (1) our approaches shed light on the deviations between designed and executed learning paths, and, (2) successful and unsuccessful learners differ considerably in the paths they follow. We believe that our work can provide instructional designers a valuable analysis tool to improve the design of both online courses and MOOC platforms in the future as they provide data-driven insights into the actual behavior of learners and the impact of their behaviors on learning outcomes.

2. RELATED WORK

In this section, we elaborate on existing research in learner modeling [5], focusing on works that investigate learning activity sequences and their impact on learning outcomes.

The problem solving behavior of learners in the context of e-learning and intelligent tutoring systems has been explored in [10, 13, 14, 19]. In contrast to our work, which considers a range of activities learners perform throughout a course (and compares them to the designed learning path), these works have explored learners' exhibited behavior within only one activity type — problem solving. Specifically, Köck and Paramythis [14] performed activity sequence clustering (an application of sequential pattern mining [22]) to model the learners' behavior, while in [13] automated clustering and human synthesis of the generated clusters were combined to identify patterns of problem solving. Shanabrook et al. [19] introduced a semi-automatic approach to identify a student's state while problem solving (including: gaming the system, guessing out of frustration, abusing hints, being on-task) in a high school-level intelligent tutoring system employing sequence-based motif discovery. Jeong and Biswas [10] developed a Hidden Markov Model to describe how different middle school student behavior trends lead to different learning processes & outcomes when problem solving.

In the context of MOOCs, sequences of learning activities have been explored by Wen and Rosé [23], who investigated the most common two-step activity sequences learners exhibit across two MOOCs. These patterns were then manually checked and analysed for interesting learning habits. A similar analysis of two-step chains was performed in Guo and Reinecke [8] who found that learners generally progress through the course content in a non-linear, "exploratory" manner [16]. Guo and Reinecke [8]'s observation of learners frequently performing "backjumps" (moving from a quiz to a lecture video previously introduced) can be considered as one of the first comparisons of executed and designed learning paths in MOOCs. Kizilcec et al. [11] (replicated in [6]) have also taken steps in this direction, by utilizing the assessment submission times (either on track, late or never) in MOOCs as indicators of learner engagement groups (completing, auditing, disengaging or sampling learners). Our work can be considered a significant expansion to these approaches, as we explore longer activity sequences (eight-step chains), thus enabling the discovery of more high-level and complex patterns and making designed vs. executed paths the focal point of our investigation.

Video interactions in MOOCs were the focus of Sinha et al. [20], who categorized the most prominent chains of video interactions (pause, play, speed, and skipping) and analyzed them with respect to learner dropout.MOOC discussion patterns have been investigated by Brooks et al. [3] who found that MOOC students exhibit markedly different discussion patterns than were expected based on blended learning environments. This finding can also be considered as a motivation for our work; MOOCs may not always be used by learners the way the instructors or course designers intended.

The concepts of process mining and conformance checking, in particular, are also employed in areas such as business process execution; [18] explains how business processes can be monitored (process mining) and then compared to the intended model (conformance checking) via a measure of fitness.

3. SUBJECTS & DATA

We explore our research question in the context of four MOOCs: `Functional Programming` (teaching the functional programming paradigm), `Data Analysis` (teaching spreadsheet and basic Python skills for data analysis), `Framing` (the art of political debates), and `Responsible Innovation` (a MOOC on the ethics and safety of new technologies). All MOOCs were offered on the edX platform in 2014/2015 and designed as xMOOCs.

Overview of MOOCs. Table 3 provides an overview of the four MOOCs in this study. The learner enrollment varies between ≈9k and ≈37k. While the four MOOCs are comparable in their video material offerings (between 41 and 59 videos), they differ significantly in the number of summative assessment questions (between 26 and 288 quiz questions). We also observe considerable differences in the percentage of video material watched by certificate-earning learners (replicating [8]) — less than half of the videos are accessed by successful learners in `Data Analysis`, while more than two thirds of the videos are accessed by successful learners in `Functional Programming`. Lastly, we note that the `Responsible Innovation` MOOC is an outlier with respect to the percentage of learners that passed the course *without* streaming any video material,[2] with nearly 20% of successful learners falling into this category; the same applies for only ≈4% of learners in the other three MOOCs.

Translating Log Traces into a Semantic Event Space. The edX platform provides a great deal of timestamped log traces, including clicks, views, quiz attempts, and forum interactions. We adapted the `MOOCdb`[3] toolkit to our needs and translated these low-level log traces into a data schema that is easily query-able.

For this work, we focus on four event types as listed in Table 2: events related to videos, quizzes, progress pages, and discussion forums. Videos can be watched - this event

[2]Note that the log traces did not capture video downloads and subsequent offline watching.

[3]`http://moocdb.csail.mit.edu/`

MOOC	Enrolled	Pass Rate	Chains Pass/Non-p.	Weeks	Videos	Quiz Questions	Passing Grade	Tries	Videos Accessed	Missing
`Functional Programming`	37,485	5.3%	1.06M/807k	14	41	288	60%	1	67.5%	4.3%
`Responsible Innovation`	8,850	4.3%	66k/30k	7	47	75	59%	1-3	49.7%	19.6%
`Framing`	34,017	2.4%	95k/141k	6	55	26	50%	2	51%	3.8%
`Data Analysis`	33,515	6.5%	1.02M/855k	8	59	136	60%	2	45%	3.6%

Table 1: Overview of the MOOCs in our study. The #Chains column contains the number of events observed throughout the MOOC (cf. Table 2). The "Passing Grade" shows the percentage of quiz questions to answer correctly to receive a course certificate. "Tries" indicates how many attempts a learner has per question. "Videos Accessed" shows the average % of course videos watched by certificate-earning learners. "Missing" is the % of certificate-earning learners who streamed zero video lectures.

Video	Quiz	Progress	Forum
`WATCH`	`START`	`VIEW`	`START`
	`SUBMIT`		`SUBMIT`
	`END`		`END`

Table 2: Overview of events considered in this work.

is generated whenever a user clicks the video 'play' button. Quizzes are identified through the beginning of the quiz session (the user enters the quiz page), the submission of one or more answers[4], and the ending of the quiz session (the user leaves the quiz page). Those quizzes are typically summative in nature. If a user views his or her progress page, the `VIEW` event is elicited. Finally, we condense discussion forum events into three kinds of items: the start of a forum session (the user first enters the forum), the submission of content (question, comment or reply) and the end of the forum session (the user leaves the forum page).
All *executed* learning paths that we extract from the learner log traces consist of the events listed in Table 2. The rationale for choosing these events comes from the designed learning path by which xMOOCs are typically formed: first watch one or more lecture videos, and then move on towards the quiz and/or forum section for assessment and knowledge building & verification respectively. In Figure 2 we visualize a week's designed learning path for each of the four MOOCs we study (this pattern is repeated in every course week). Video lectures form a common denominator, starting the path. `Functional Programming` and `Data Analysis` rely on videos and quizzes only (with `Data Analysis` exhibiting multiple video-quiz "cycles" within a week), whereas `Responsible Innovation` and `Framing` make use of the forums as well. The learning path shown for `Framing` does not include quizzes as they are posed only in the final week (in the form of an exam).

4. APPROACH

Having introduced the subjects of our work and the events we consider, we now describe the three distinct approaches to the visualization & exploration of executed learning paths (that is, learners' sequential movement *over time* through the activities offered in a MOOC) we developed.

[4]Note that on the edX platform answers to individual quiz questions are submitted (instead of all answers at once).

4.1 Video Interactions

As shown in Figure 2, videos are a focal point of xMOOCs. Accordingly, in a first analysis, we focus exclusively on video interactions and explore to what extent learners adhere to the designed video watching learning path. Therefore, in this study we only make use of `WATCH` events.
We transform the `WATCH` events generated by a set of learners L across the duration of a MOOC $\mathcal{M}$ into a directed graph $G_{\mathcal{M},L} = (V_{\mathcal{M}}, E_{\mathcal{M},L})$ — as the subscripts indicate, with $\mathcal{M}$ fixed, the set V is independent of the subset of learners chosen, while E is dependent on the learners in L. All lecture videos contained in $\mathcal{M}$ form the set of vertices $V_{\mathcal{M}}$. The vertices are labelled chronologically, that is, for any vertex pair (v_i, v_j) with $i < j$, the corresponding lecture video i must appear in the designed learning path before video j.
The edges are directed and weighted according to the number of `WATCH` events by the learners L: an edge between v_{i-1} (source) and v_i (target) presents the learners' transition between these videos, i.e. the number of times learners watching video v_{i-1} watch v_i next, before any other video. We disregard self-loops (watching the same video again) as we are focusing on the progression of the learners through the set of lecture videos.
Having generated $G_{\mathcal{M},L}$, we now turn to its visualization (to aid instructors and course designers): the vertex layout is sequential and governed by the designed learning path through the videos (represented as vertices). For MOOCs with thousands of participants it is likely that every single video pair combination possible is contained in at least one learning path. To avoid visual clutter, we filter out the most *infrequent* edges: we bin the edges according to the week their source vertex appears in and remove the 10% of edges that occur most infrequently in this course week.
To discover whether or not there are marked differences in the way different groups of learners behave, we generate the video interaction graph for different sets of learners, such as successful (certificate earning) vs. unsuccessful learners.

4.2 Behavior Pattern Chains

Having considered the transitions between lecture videos, we now turn to the exploration of transition patterns among all eight events identified in Table 2. Previous works [23] have viewed MOOC learner patterns either in terms of one-step directed pairs of events (such as *watch video* → *begin quiz*) or based on video click chains only [20].
One-step chains can only provide limited insights into more high-level behavioral patterns — we may, for instance, be in-

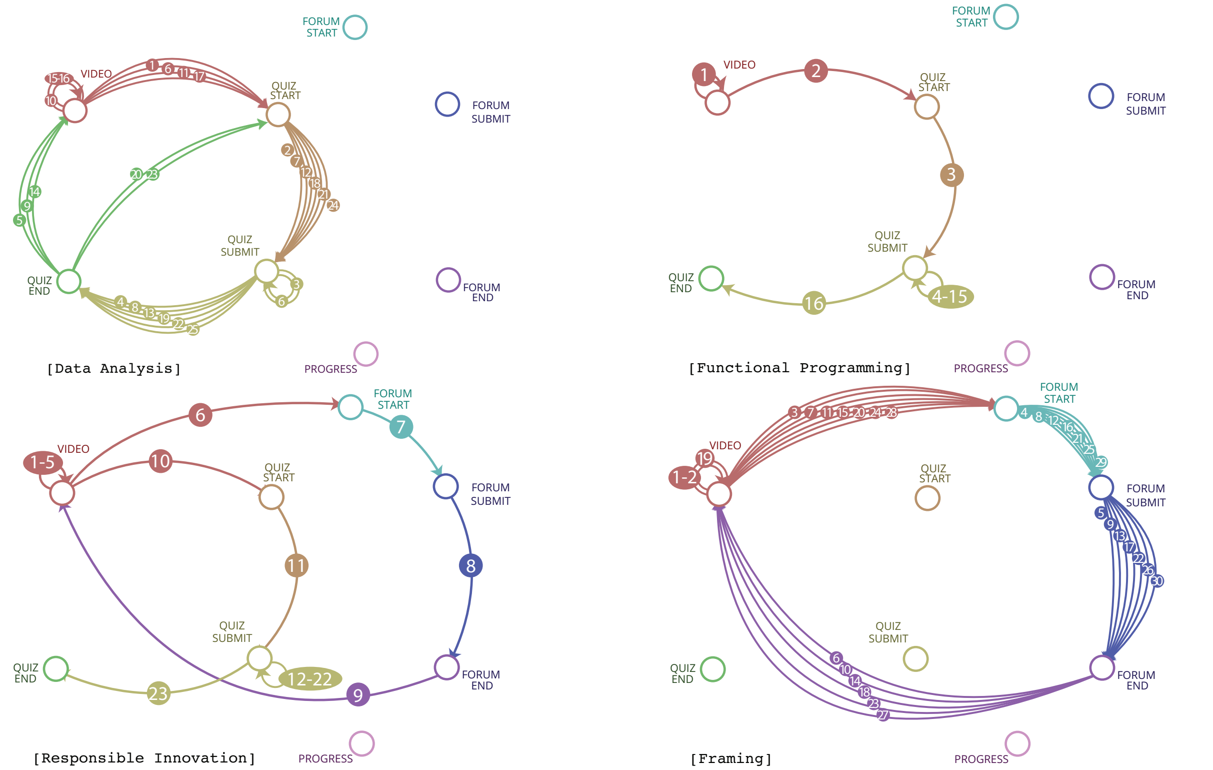

Figure 2: The designed learning path for a standard week (Week 4) of each MOOC. The circled numbers indicate the step number of each transition in that week's sequence. Notice the diversity in course designs that characterize these four MOOCs.

$\texttt{QUIZ}_{START} \rightarrow \texttt{QUIZ}_{END} \rightarrow \texttt{WATCH} \rightarrow \texttt{WATCH}$

$\rightarrow \texttt{WATCH} \rightarrow \texttt{WATCH} \rightarrow \texttt{WATCH} \rightarrow \texttt{WATCH}$

Figure 1: An example eight-step chain.

terested to understand how many learners are "binge watchers" (watching many videos in a row) or "strategic learners" (looking at quiz questions before watching the corresponding lecture video). In order to contribute insights to our research question we need to consider longer chains. We have settled on eight-step chains, as they provide insights into more high-level concepts but are still numerous enough in our log traces to make claims about their general usage. We consider all events of Table 2 and create event chains by sliding a window of size eight over each learner's chronologically ordered learning path through a MOOC. An example eight-step chain this procedure yields is shown in Figure 1.
To identify the underlying trends in the chains, we employed the *open card sort* approach [7]. After printing out two sets of the thirty most frequently occurring chains on paper, two authors independently sorted them into (non-predefined) like-groups by hand and afterwards discuss the differences in each sort, creating a composite of the two results. The outcome of this method is a synthesis of similar chain types into groups sharing the same *motif*, or recurring theme. Based on the motifs, we created a rule-based system that assigned a MOOC's entire set of chains to the identified motifs (chains that do not fit into any motif are left "unassigned"). This process is repeated for each of the MOOCs we investigate. The advantage of this approach over the automatic clustering of the chains is the infusion of our domain knowledge into the clustering process.

4.3 Event Type Transitions

Lastly, we explore event type transitions, or how likely learners are to move from one event type to another. Inspired by the methods employed in [10, 13, 14] we use discrete-time Markov chains (a memory-less state transitioning process encoding how often learners move from one event type to another) in order to chart the likelihood that a learner will transition from one engagement activity to another. Whereas the prior works employ these methods in the context of problem solving (knowledge assessment), we focus on the larger process of *knowledge building*, which transpires over the span of an entire course.
While it may be self-evident that non-passing learners answer less quiz questions than their certificate-earning peers (and thus the transition probabilities to $\texttt{SUBMIT}_{QUIZ}$ are likely to be lower for non-passers), the visualization of the Markov chains enables designers to pinpoint the differences in transitions between different types of learners (e.g. passers vs. non-passers) across all events in one coherent plot.

5. FINDINGS

To answer our research question (do learners adhere to the designed learning path?), we apply the three approaches outlined in Section 4 to the datasets described in Section 3.

5.1 Video Interactions

We visualize the video interactions across the first three weeks (these are where the most deviations occur; the later weeks are more in line with the designed path) of each MOOC in Figures 3 to 6, distinguishing two sets of learners: those that eventually earn a certificate ("Passing") and those that do not ("Non-Passing"). The *designed* video interaction learning path is exhibited by the left-to-right flow of the vertices (one per video). The edges correspond to the *executed* learning paths — with edge thickness indicating the (normalized) number of learners having taken that path (only the 90% most frequently occurring transitions each week are shown); the set of red edges represent the executed transitions that follow the designed transitions. A number of observations can be made based on the visualizations: (i) passing learners deviate considerably less from the designed learning path than non-passing learners across all four MOOCs, (ii) passing learners are more likely to skip video lectures introducing the platform (the first three videos in the `Framing` MOOC) than non-passing learners, indicating a higher level of seniority in MOOC-taking, (iii) towards the end of week three, the deviations among the sets of passing and non-passing learners are negligible (i.e. the non-passing learners still active exhibit a similar video watching behavior as the passers), and (iv) skipping videos — jumping ahead — is much more common than backtracking — jumping backwards — for both passers and non-passers.

An emerging object in the field of Design (and gaining some attention in the field of Software Design [4]) is that of *desire paths*, or paths not intended by the designer, but those which "arise due to off-[path] use ... for a variety of purposes such as access to places of interest and shortcutting" [2]. This research serves as a reminder that desire paths indeed exist in MOOCs (as evident in the skipping of introductory lecture material) — they just have not yet been made as visible as those brown stripes of beaten grass and dirt transecting public parks and trails. They are a reminder that humans can collectively communicate good design by their actions.

5.2 Behavior Pattern Chains

Our second approach explores learners' behavioral patterns. As outlined in Section 4.2, we first manually clustered and labelled the most frequent eight-step pattern chains in order to determine what type of behaviors (or motifs) learners exhibit beyond a single-click transition, before automatically assigning the remaining chains into those motifs. Depending on the MOOC, this approach yielded between eight and 11 motifs, with some motifs appearing only in a subset of courses. For brevity reasons, in Tables 3 to 6 for each MOOC we list its most frequent motifs (specifically those into which ≥2% of all chains are classified); as a comparison in Table 3 we also list the total number of chains generated by passing/non-passing learners in each MOOC — depending on the MOOC, the listed motifs capture between 42%–77% of the total number of chains. Whenever a motif is first introduced, we briefly describe which event types and event

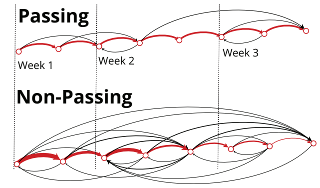

Figure 3: `Functional Programming` video interactions.

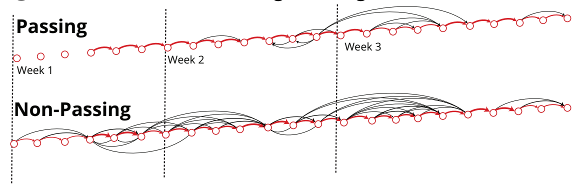

Figure 4: `Framing` video interactions.

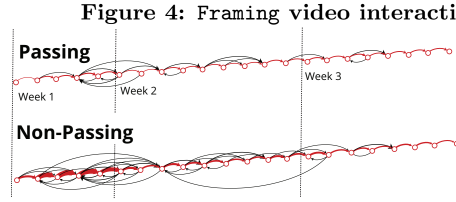

Figure 5: `Data Analysis` video interactions.

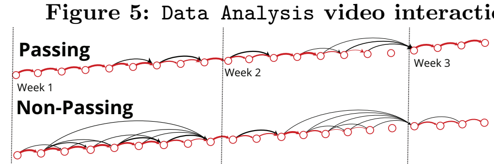

Figure 6: `Responsible Innovation` video interactions.

orderings characterize it[5].

Examining the results, we observe that (i) *Binge Watching* is a frequent motif in all MOOCs with non-passers always exhibiting more binge watching (i.e. watching videos uninterrupted by other activities) than passers, (ii) the *Lecture→Quiz Complete* motif, which captures the "classic" xMOOC idea of video watching with subsequent question answering is frequent in three of the four MOOCs[6], however no consistent divergent behavior for passers and non-passers is found, (iii) motifs with forum events occur in three of the four MOOCs — by course design in `Framing` and `Responsible Innovation` (cf. Figure 2), but not in `Functional Programming`, indicating issues related to material clarity, and (iv) the *Quiz Check* motif, which is exhibited by learners checking the quiz questions without answering any of them (which is usually followed by video watching and subsequent quiz completion), is only found in one MOOC frequently; in `Data Analysis` 2% of the chains follow this motif, a smaller percentage than we expected, indicating that very few learners are gaming the system by "attempting to succeed in an educational environment by exploiting properties (quiz ques-

[5]Note, that we implemented our rules for the automatic assignment of chains to motifs according to these characterizations.

[6]It does not appear among the frequent motifs in `Framing`, which has a final exam instead of weekly quizzes.

tions are posted alongside the video material) of the system (edX platform) rather than by learning the material and trying to use that knowledge to answer correctly," [1].

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	552,363 (29.4%)	328,995 (30.8%)	223,368 (27.7%)
	X_{QUIZ} events only with at least one X = *SUBMIT*			
2	Binge Watching	149,784 (8%)	59,498 (5.6%)	90,286 (11.2%)
	`WATCH` events only			
3	Lecture→Quiz Complete	100,179 (5.3%)	50,415 (4.7%)	49,764 (6.2%)
	`WATCH` event(s) followed by X_{QUIZ} events; at least one X = *SUBMIT*			
4	Quiz Complete→Forum	99,828 (5.3%)	67,722 (6.3%)	32,106 (4%)
	X_{QUIZ} events (at least one X = *SUBMIT*) followed by X_{FORUM} events			
5	Quiz Complete→Progress	38,854 (2.1%)	26,126 (2.4%)	12,728 (1.6%)
	X_{QUIZ} events (at least one X = *SUBMIT*) followed by $X_{Progress}$ events			

Table 3: Most frequent motifs (≥2% chains) in `Functional Programming`.

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	18,446 (16.6%)	11,377 (14.7%)	7,069 (21.1%)
2	Binge Watching	12,530 (11.3%)	8,461 (10.9%)	4,069 (12.1%)
3	Lecture→Quiz Complete	5,060 (4.6%)	3,752 (4.8%)	1,308 (3.9%)
4	Lecture→Forum→Lecture	3,910 (3.5%)	2,386 (3.1%)	1,524 (4.5%)
	`WATCH` events followed by X_{FORUM} events followed `WATCH` events			
5	Quiz Complete→Progress	3,741 (3.4%)	2,898 (3.7%)	843 (2.5%)
6	Quiz Complete → Lecture → Quiz Complete	2,277 (2.1%)	2,019 (2.6%)	258 (0.8%)

Table 4: Most frequent motifs (≥2% chains) in `Responsible Innovation`.

5.3 Event Type Transitions

The Markov models of our four MOOCs are visualized in Figures 7 to 10. Since we observe the same event types across the four MOOCs, the set of vertices, their placement in the visualization, and their semantics are identical. To minimize visual clutter, we only plot the transitions (i.e. the edges) that exhibit a probability of 0.2 or higher. Once more we make the distinction between passing and non-passing learners. The resulting visualizations show the behavioral differences not only between passing and failing students within a given course, but these also allow for cross-course analyses which shed light on what types of behavioral patterns define a course. For example, when comparing `Framing` (Figure 9) and `Data Analysis` (Figure 7), marked differences in their pedagogical structure are evident; `Framing` appears to foster a very social, collaborative environment, whereas `Data Analysis` learners mostly focus their attention on lectures and assessments, with little concern for discussion. The visualizations also reveal at which specific moments learners seek feedback on their progress (i.e. make a transition to the Progress vertex), such as after a Quiz or Forum in `Responsible Innovation` and `Framing`. These movements are *not* included in any of the courses' designed paths; course designers can use this insight to proactively insert feedback in order to encourage more awareness and self-regulated learning. When comparing transitions of passing vs. non-passing learners, we observe that (i) non-passers make the transition to the video event from more diverse event types than passers (indicating that non-passers' executed paths follow the designed path to a lesser degree than passers' executed paths), (ii) video-to-video transitions are more prevalent among non-passers (in line with our findings on the binge watching motif), and (iii) passing learners are more likely to move from *Quiz Start* to *Quiz Submit*, while non-passing learners are more likely to move from *Quiz Start* to *Quiz End* (without answering a question).

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Binge Watching	64,822 (27.3%)	18,023 (18.9%)	46,726 (33.1%)
2	Lecture→Forum→Lecture	29,224 (12.3%)	11,651 (12.2%)	17,505 (12.4%)
3	Quiz Complete	12,984 (5.5%)	9,156 (9.6%)	3,781 (2.7%)
4	Forum→Lecture	7,850 (3.3%)	3,035 (3.2%)	4,800 (3.4%)
	X_{FORUM} events followed `WATCH` events			
5	Lecture→Forum	7,488 (3.2%)	3,008 (3.2%)	4,462 (3.2%)
6	Quiz Complete→Lecture→Quiz Complete	5,551 (2.3%)	4,022 (4.2%)	1,501 (1.1%)

Table 5: Most frequent motifs (≥2% chains) in `Framing`.

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	169,786 (9%)	116,878 (11.4%)	52,908 (6.2%)
2	Quiz Complete→Lecture→Quiz Complete	145,596 (7.7%)	82,247 (8%)	63,349 (7.4%)
3	Binge Watching	87,760 (4.7%)	28,066 (2.7%)	59,694 (7%)
4	Lecture→Quiz Complete	78,790 (4.2%)	41,543 (4.0%)	37,247 (4.4%)
5	Quiz Complete→Lecture	43,612 (2.3%)	21,916 (2.1%)	21,696 (2.5%)
6	Quiz Check	37,406 (2%)	19,444 (1.9%)	17,962 (2.1%)
	$QUIZ_{START}$ followed by $QUIZ_{END}$ events			

Table 6: Most frequent motifs (≥2% chains) in `Data Analysis`.

6. CONCLUSION

Before adaptive learning systems can reach their potential, two important baselines must be established: (i) the precise learning path the instructor wants the student to follow and (ii) students' natural behavior within the course. Adaptive instruction will be most effective when the differences between these two baselines are both identified and addressed. The present research offers novel insights into the identification of those differences.

Specifically, in this work we have introduced three different approaches (the video interaction graph, behavior pattern chains and event type transitions) to explore and visualize MOOC log traces with respect to the designed and executed learning paths.

We have applied our approaches on the log traces of four different edX-based MOOCs (from different domains and different pedagogical structures) and have shown to what extent learners (as a whole group as well as partitioned into passing and non-passing learners) follow the prescribed path.

In future work, we will expand our analyses to a larger set of MOOCs to gain a greater understanding of the "classes" of xMOOCs that exist on the major MOOC platforms today. We also plan to consider more diverse sub-populations of learners in future analyses, beyond passing or not passing. We will also investigate semi-automatic approaches to the adaptation of MOOC learning paths, in order to minimize the gap between designed and executed paths as well as the impact this work has on engagement, retention, learner success and more fine-grained learner partitions (such as completing, auditing, and sampling learners [11]).

References

[1] Ryan S.J. Baker, Albert T. Corbett, Kenneth R. Koedinger, and Shelley Evenson *et al.* Adapting to when students game an intelligent tutoring system. In *Intelligent tutoring systems*, pages 392–401, 2006.

[2] Lori EA Bradford and Norman McIntyre. Off the beaten track: Messages as a means of reducing social trail use at St. Lawrence Islands National Park. *Journal of Park and Recreation Administration*, 25(1):1–21, 2007.

[3] Christopher Brooks, Jim Greer, and Carl Gutwin. The data-assisted approach to building intelligent technology-enhanced learning environments. In *Learning Analytics*, pages 123–156. 2014.

[4] Christian Crumlish and Erin Malone. *Designing social interfaces: Principles, patterns, and practices for improving the user experience.* O'Reilly Media, Inc., 2009.

[5] Michel C Desmarais and Ryan SJ d Baker. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1-2):9–38, 2012.

[6] Rebecca Ferguson and Doug Clow. Consistent Commitment: Patterns of Engagement across Time in Massive Open Online Courses (MOOCs). *Journal of Learning Analytics*, 2(3):55–80, 2016.

[7] Sally Fincher and Josh Tenenberg. Making sense of card sorting data. *Expert Systems*, 22(3):89–93, 2005.

[8] Philip J. Guo and Katharina Reinecke. Demographic differences in how students navigate through MOOCs. In *L@S 2014*, pages 21–30, 2014.

[9] Sherif Halawa, Daniel Greene, and John Mitchell. Dropout prediction in MOOCs using learner activity features. *Experiences and best practices in and around MOOCs*, 2014.

[10] Hogyeong Jeong and Gautam Biswas. Mining student behavior models in learning-by-teaching environments. In *EDM*, pages 127–136, 2008.

[11] René F Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *LAK 2013*, pages 170–179, 2013.

[12] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. Predicting MOOC dropout over weeks using machine learning methods. In *EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65, 2014.

[13] M. Köck and A. Paramythis. Towards Adaptive Learning Support on the Basis of Behavioural Patterns in Learning Activity Sequences. In *Intelligent Networking and INCOS '10*, pages 100–107, 2010.

[14] Mirjam Köck and Alexandros Paramythis. Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction*, 21(1-2):51–97, 2011.

[15] Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. Retention and intention in massive open online courses. *Educause Review*, 48(3):62–63, 2013.

[16] Jens O Liegle and Thomas N Janicki. The effect of learning styles on the navigation needs of Web-based learners. *Computers in Human Behavior*, 22(5):885–898, 2006.

[17] Daniel FO Onah, Jane Sinclair, and Russell Boyatt. Dropout rates of massive open online courses: behavioural patterns. *EDULEARN14*, pages 5825–5834, 2014.

[18] Anne Rozinat and Wil MP van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.

[19] David H Shanabrook, David G Cooper, Beverly Park Woolf, and Ivon Arroyo. Identifying high-level student behavior using sequence-based motif discovery. In *EDM*, pages 191–200, 2010.

[20] Tanmay Sinha, Patrick Jermann, Nan Li, and Pierre Dillenbourg. Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions. *EMNLP 2014*, pages 3–14, 2014.

[21] Sarah D. Sparks. Data Mining' Gains Traction in Education. *Education Week*, 2010.

[22] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explor. Newsl.*, pages 12–23, 2000.

[23] Miaomiao Wen and Carolyn Penstein Rosé. Identifying latent study habits by mining learner behavior patterns in massive open online courses. In *CIKM '14*, pages 1983–1986, 2014.

[24] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rose. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *2013 NIPS Data-driven education workshop*, volume 11, 2013.

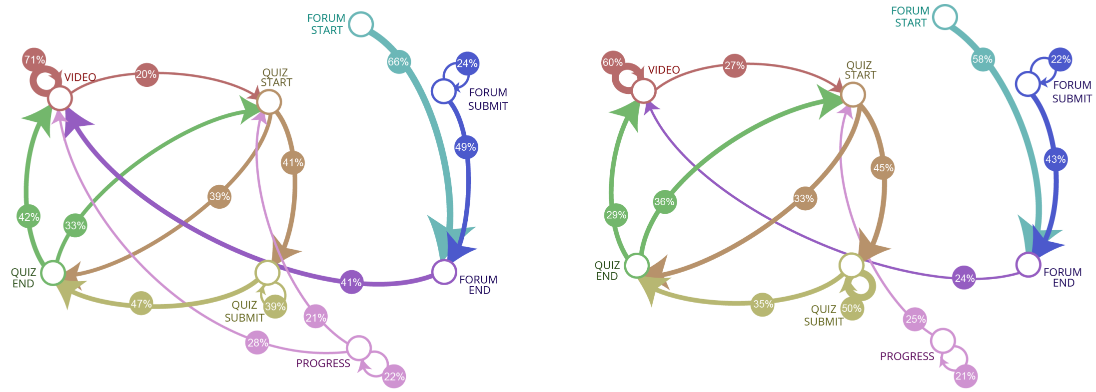

Figure 7: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Data Analysis` MOOC. Edges with weights below 20% are hidden from view.

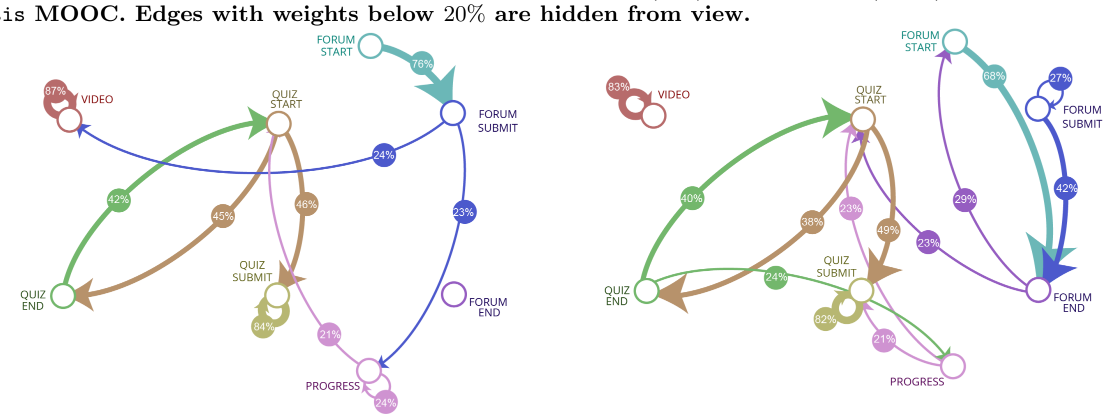

Figure 8: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Functional Programming` MOOC. Edges with weights below 20% are hidden from view.

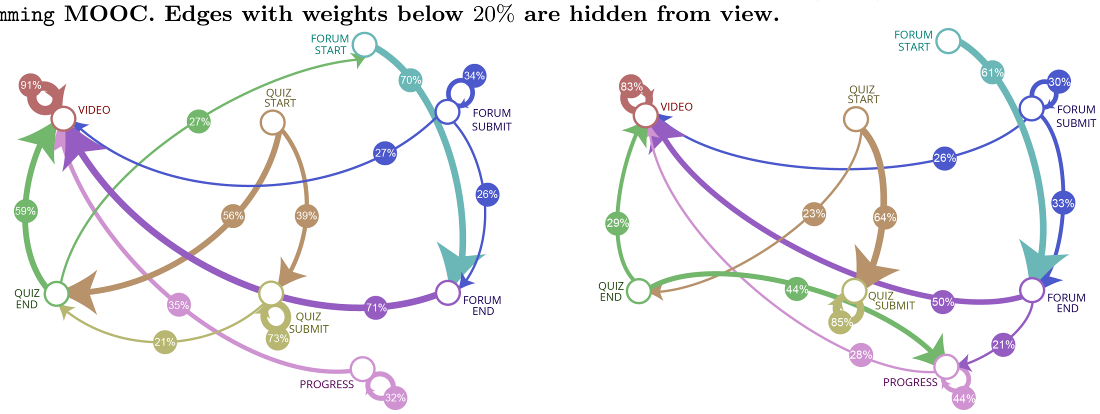

Figure 9: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Framing` MOOC. Edges with weights below 20% are hidden from view.

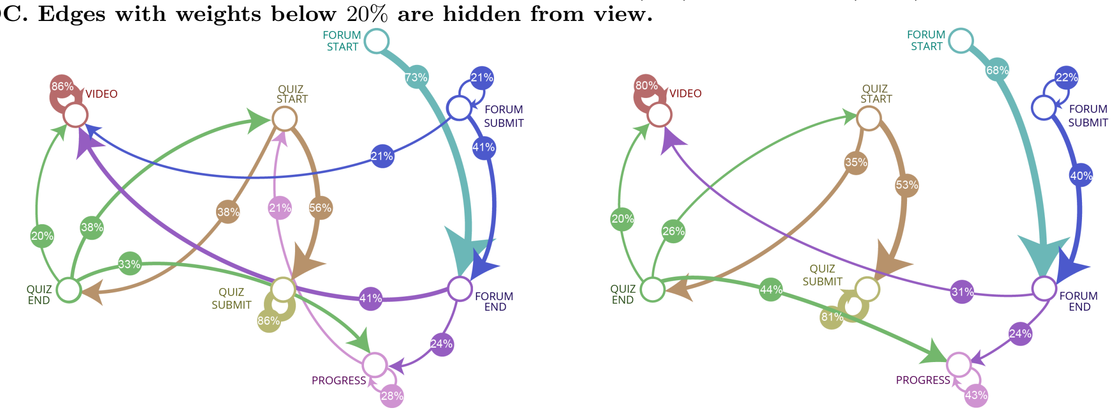

Figure 10: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Respon-sible Innovation` MOOC. Edges with weights below 20% are hidden from view.