Analyzing Student Procrastination in MOOCs: A Multivariate Hawkes Approach

Mengfan Yao
Department of Computer Science
University at Albany - SUNY
myao@albany.edu

Shaghayegh Sahebi
Department of Computer Science
University at Albany - SUNY
ssahebi@albany.edu

Reza Feyzi Behnagh
Department of Educational Theory & Practice
University at Albany - SUNY
rfeyzibehnagh@albany.edu

ABSTRACT

Student procrastination, as the voluntary delay of intended work despite expecting to be worse off for the delay, is an important factor with potentially negative consequences in student well-being and learning. In online educational settings such as Massive Open Online Courses (MOOCs), the effect of procrastination is considered to be even more prevalent and detrimental, as online courses are often self-paced and self-directed, where higher levels of self-regulated learning are expected from the students. Past research has mainly described students' procrastination by either static time-related measures (e.g. averaged starting time over all assignments per student), or by temporal models' parameters, under the assumptions that student activities take place at a constant rate (e.g. Homogeneous Poisson models), and that student interactions with one learning material are independent of interactions with another. In this work, we propose to consider the interdependence between the students' temporal activities while modeling their sequences in a continuous time scale. To this end, we propose to model the interaction sequence between each student and each course module, i.e. each module-student pair, as Multi-dimensional Hawkes processes, which not only capture the relationship between students' learning activities and their exogenous stimuli such as assignment deadlines, but also capture the endogenous responses *within* and *between* types of learning materials. Our experiments show that not only there exists dependencies between students' historical activities and the future ones when different types of learning materials are involved, such dependencies also provide meaningful interpretations in terms of students' procrastination behaviors. Furthermore, our findings show that in addition to association with delay, the parameters learned by multi-dimensional Hawkes processes provide more procrastination-related information and can improve our explanation of student grades.

Keywords

Procrastination, MOOCs, Student Modeling, Multivariate Hawkes Process, Clickstream Data

Mengfan Yao, Shaghayegh Sahebi and Reza Feyzi Behnagh "Analyzing Student Procrastination in MOOCs: A Multivariate Hawkes Approach" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020),* Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 280 - 291

1. INTRODUCTION

Student academic procrastination has shown to have negative effects on students' learning and well-being. Procrastination is prevalent in different academic settings like traditional classrooms, and could be even more widespread in online learning environments, as higher levels of time-management and self-regulated learning (SRL) skills are required [47, 3, 16]. To describe and measure student procrastination, past research has been mainly relying on either self-reported surveys (e.g. [27]) or time-related features that are associated with students dilatory behaviors (e.g. [10]). As procrastination is inherently subjective, self-reported surveys have been heavily used in earlier research, to differentiate procrastinators and non-procrastinators by emphasizing on measuring the perceptions of the students. Although self-report survey measures capture students' retrospective reports of their studying and delaying behaviors, they are administered in a cross-sectional manner, rely on students' memory, and are usually static point estimates that summarize students' *average* degree of procrastination.

Considering the noises of self-reported data [37], in more recent studies, more focus has been given to the behavioral side of the procrastination, where time-related measures were proposed and used as the representation of students' procrastination. For example, measures such as students' average delays in starting coursework, the average time they spent in doing assignments, students' average paces of viewing lectures have been studied as factors of procrastination [2, 10, 15, 6, 22]. However, these measures lack the ability to describe students' continuous behaviors within a period of time. An analogy to such methods is to describe the entire distribution using the sample mean, without fully knowing the distribution. To tackle this limitation, more recently, emphasis has been on modeling the time points of student activities that are extracted from students' learning trajectory data (e.g. log or click-streams of student historical actions), via stochastic models. For example in [33], Park et al. modeled students' per-day activity counts during each week of the course via a Poisson mixture model, which models the entire trajectory of each student activities during a weekly module. Other factors that have been considered to be important in describing procrastination in the past research are the effects of different learning materials (e.g. forums and quizzes) as well as students' interactions with them (e.g.[1, 28]). However, to the best of our knowledge, no past work has considered the possible time dependencies within and between students' interactions with different learning material types. For example, viewing video lectures more intensively *mostly before* the first attempt of an assignment may suggest that a student prefer to learn the materials first before trying the assignment. On the other hand, watching lecture videos *dominantly after* the first attempt of an assignment may suggest that the student prefer to try the assignment first and then go through the video lectures if they encountered any problems.

To summarize, past research has attempted to describe procrastination using static time measures, or measures summarized from more sophisticated temporal models, based on students' interactions with one or more learning materials. However, two important factors of student behaviors and their association with procrastination have not been fully explored: (1) the dependencies between students' past and future interactions within each learning material type (e.g. knowing a student has looked at lecture slides at some time, how and when are they going to have the next activity?) and (2) the dependencies between students' interactions with different types of learning materials (e.g. are watching video lecture usually followed by a submission of an assignment?) In this work, we aim to address these two factors by answering the following questions: within each learning module, that is the unit of a course that learning materials are provided, (Q1) are the past activities independent of future ones? Or some activities can trigger other ones to arrive within a short period of time (i.e. time dependencies between activities)? And (Q2), are students' interactions with one type of learning material (e.g., video lectures) independent from another type (e.g., discussion forums)? Furthermore, (Q3) if such dependencies exist, how are they associated with student procrastination? (i.e. the dependencies between a student's past and future activities as well as dependencies a student's interactions with one learning material with another.)

As a result, our goal is to find the missing link between students' procrastination and students' activities within and between different types of online learning materials. To achieve this goal, we propose to use multi dimensional Hawkes processes as a powerful tool that addresses the above mentioned concerns in student procrastination analysis. Particularly, we represent all activities on one type of learning material as one dimension in the multi-dimensional Hawkes model. We show that this model better fits our data, in comparison to baseline temporal processes. Also, to answer Q1 and Q2, we demonstrate that it can capture both students' reactions to the deadlines as action-triggering factors that come externally (i.e. exogenous stimuli), and students' responses to the previous interactions with different types of learning materials, such as video lectures, assignments, and discussions (i.e. self-excitement). By doing so, we can understand students' procrastination behavior from a stochastic process point of view, with two main stimuli: (1) some of the students' activities can be viewed as a response to an external stimulus, e.g. deadlines of the assignments (2) some other student activities can be viewed as the results of previous interactions that the student had with the same or other learning material types. Based on the model parameters, to answer Q3, we also propose a measure that not only describes student procrastination but also is able to explain student performance better than the static delay measure.

The outline of this paper can be summarized as follows: In Section 2, we go over three main bodies of the related work; in Section 4 we go over the details of the dataset that we use; in Section 4, we provide the intuition of using the Hawkes model, then statistically and visually show that a Hawkes process is a proper choice for modeling module-student interactions; in Section 5, we formally define our problem and introduce the multi-dimensional Hawkes model that we use in this study. We perform various experiments in section 6, to analyze the model parameters, explain their interpretation, and associate them with procrastination as well as students' assignment grades. Finally, the conclusion of this work is summarized in Section 7.

2. RELATED WORK

Students' procrastination In the past research on student procrastination, the main focus has been on the measures that capture either students' perceptions (e.g. self-reported surveys on procrastination [35, 41, 40, 12, 23]), or static measures that describe students' dilatory behaviors as the representation of procrastination [15, 11, 44, 33]. For example, in [10], Cerezo et al. studied 140 undergraduate and used measures such as students' delay and time-spent variables to describe procrastination. For another example, in [2], Asarta and Schmidt studied students' behaviors in accessing lecture notes of a blended-learning course, and proposed to use features such as pacing, anti-cramming, and consistency in reviewing course materials. A few recent works have tried to model student activities to provide a temporal perspective of procrastination behavior. For example, Backhage et al. proposed a model that captures procrastination-deadline cycles of all students in the course using a stochastic temporal model [4]. However, this model assumes that all students follow the same procrastination behavior during the course and does not distinguish the differences between student behaviors. In [33], Park et al. assumed that students' daily activity counts follow a mixture Poisson distribution, which is a mixture of a procrastination component and a non-procrastination component. Particularly, by assuming the independence between students' past and future activities, they proposed to model each day of the week by a Poisson with a constant rate for all weeks. In the end, they described procrastinators as the ones with a dominant procrastination component versus the non-procrastination one, i.e. the students who have a fast-increasing activity counts towards the end of the week. Moon et al. assumed that procrastination behavior over time can be described by a curvilinear growth curve and modeled it using latent growth curve modeling. To validate this assumption, they compared the curvilinear model with a non-growth and a linear model and showed that their model has a better goodness-of-fit than the baseline models [30]. In contrast with the existing research that either uses summary variables or ignores the dependence between different student activities, in this paper we aim to model the temporal activity interrelationships and associate them with student procrastination.

Modeling students' engagement using their learning trajectories. Other relevant studies to our work are the ones that model student learning trajectories to understand other aspects of their behaviors, such as student engagement in online learning environments. While many past studies focused mostly on utilizing cumulative factors such as frequency of watching videos or using discussion forums [39, 17, 34, 13], more recent work attempted to build more complex models of student behaviors. For example, in [46], Zhu et al. constructed students' social connection networks based on students' weekly post-reply dynamics, along with node attributes, such as assignment scores. Particularly, they used an exponential random graph model to compute the structural features of the social connection networks, to understand the relationship between students' engagement in the forums and their performances in the assignments. In another example, Lan et al. proposed a statistical model, which consists of two components: a learning model and a response model [25]. These two models represent nine behavioral features extracted from students' video-watching clickstreams and in-video quiz responses in one MOOC course, with the aim to find the behavioral features that lead to high levels of student engagement. Similarly, Kizilcec et al. classified students' behaviors based on binary features extracted from students' log data (1 if a student had any activities that are associated with a learning material, 0 otherwise, for all learning materials) [24]. As a result, they identified 4 behavioral types: completing, auditing, disengaging, and sampling, from these binary features. For another similar example, Gelman et al. extracted student features from students' log data as well and applied Nonnegative Matrix Factorization to find 5 types of student behaviors - deep, consistent, bursty, introduction, and sampling [18]. Particularly, the authors used a procrastination indicator as a feature, that is, the average amount of time left before the deadline when a student submits their assignments. In summary, past research on modeling student engagement is similar to our study in the sense that the models utilize students' activities that were extracted from students' log data. However, it differs to ours in the following two ways: (1) the models usually define students' engagement levels based on the *counts* of students' historical actives, without directly modeling students' learning trajectories as stochastic processes, (2) the aim is usually to model or predict students' future engagement levels, rather than studying students' procrastination and its association with students' performance.

Hawkes in education. Hawkes processes, a family of stochastic point processes, have been frequently used to model complicated time-stamped events in continuous time. Due to Hawkes process's capability to model scenarios where historical events influence future activities, it has been frequently used in finance [5] and seismology [32] and has been gradually becoming a useful modeling tool in the domain of social media [7, 36, 29], as well as recommendation systems [14, 43, 21, 38]. In the education domain, a few works have used Hawkes processes so far, especially to model social and interaction data among students [19, 20, 26]. For example, Lan et al. proposed a single-dimensional Hawkes model to recommend relevant discussion threads to students according to their historical interactions with course forums. In a similar application, a Hawkes model is suggested by Von Davier et al. to model the collaboration dynamics between students within and between groups [42]. Along this line, Halpin et al. used multi-dimensional Hawkes processes to understand students' collaboration with each other [19, 20]. Another interesting application of the Hawkes process in the education domain is the work by Boerner et al. that analyzed the association between student skills and the skills required by professional jobs [8]. In another recent work, Cai et al. used Hawkes processes as a step in their model to predict which video a student will watch next based on their historical interactions with the videos in an edX course [9]. They use long and short multi-dimensional Hawkes processes that differentiates the long-term and short-term temporal dependencies between video-watching actions. None of the above works uses the Hawkes processes to model the procrastination behavior, nor considers course deadlines and milestones in their application of the Hawkes process.

3. DATASET

Our dataset is publicly collected from the Canvas Network[1] MOOC platform [31], which is an online platform that hosts various open online courses in different academic disciplines, such as Computer science, Social Science, and Business management. These courses have multiple types of learning resources, including Wiki pages, assignments (or quizzes), and discussions. Assignments can be quiz-style or in a longer format where students need to upload a file to complete the submission. Each learning module is associated with one Wiki page. In total, CANVAS data contains 389 anonymized courses where the names of students and courses along with the contents of discussions and assignment (or quiz) submissions are not available.

In this work, we mainly focus on exploring the student learning trace data. Specifically, we select a computer science course (course id: 770000832960058) that best fits the following criteria: (1) having a large number of students[2]; (2) including multiple types of learning materials (such as video lectures, assignments, discussions); and (3) containing a large number of student historical learning activities. To obtain student learning activity data, we use Canvas logs files (Pageview requests). We divide the learning activities into three types. Specifically, we consider viewing the lectures, downloading the attached files, and previewing the attached files as the activities associated with video lectures (L). Activities that include viewing, creating, saving, updating, and submitting each assignment attempt are associated with assignments (A). Finally, we consider reading (marking as read), subscribing, creating, replying, and editing discussion entries, discussion topics, and direct messages as discussion-related activities (D).

We separate the data in module-student pairs, as we aim to model each student's interactions with each individual learning module. As each module has its specific deadlines according the course design, we choose each module rather than the whole course as the unit of our study. Also, different modules usually have different learning objectives, which will possibly trigger different behaviors. By doing so, we are able to capture a finer granularity of the data. Finally, we have 731 students and $\sim$ 946K learning activities in the selected course.

[1] http://canvas.net

[2] Enrolled students who have missed more than 50% of the assignment submissions during the courses, along with those who did not receive a final grade, are considered as dropouts and are disregarded in this study.

4. BACKGROUND: HAWKES PROCESS AS A FIT TO STUDENT ACTIVITIES

Since we want to study the interactions between students and modules from a temporal aspect, point processes are one of the best choices for our application. Additionally, because of the interaction irregularities in our application, we must select a point process that can handle this type of information. Specifically, past studies have shown that students' activities can take place in an irregular manner during various periods of a course, particularly affected by milestones such as assignment deadlines and exam dates [16]. As a result, the point processes that follow a constant rate, such as Poisson processes, are not the appropriate model for our application. In Poisson processes, the main assumption is the independence between past and future occurrences of the events, which can not be met in student studying behaviors. Not only some student activities happen in response to the course milestones, but also a part of these activities can be interrelated with each other. For example, a student whose goal is to start discussing a topic in the discussion forum may watch a video lecture about the same topic before posting in the forum. To meet the temporality and interdependence assumptions of our application, we choose to model student activities during course modules with Hawkes processes.

One of the most important properties of the Hawkes process is its ability to deal with the interrelationships between future and past activities. This is in contrast with the memoryless Poisson process where all activities are assumed to be independent of each other. More importantly, the Hawkes process allows the activities to be excited both exogenously (by external stimuli, similar to the Poisson process) and endogenously (self-excitement, by internal stimuli). In other words, the Hawkes process has a branching process point of view. It assumes that some activities arrive as a result of exogenous stimuli (i.e. immigrant activities). Then, the immigrant activities can trigger their following activities (i.e. offspring activities), and those offspring activities can further trigger their own offspring activities, and so on. That is, the offsprings of an immigrant activity are structured into a latent cluster because they are all triggered by the same immigrant and arrive more closely to each other than the activities that are in other clusters.

As a result, the Hawkes process can capture more information than the Poisson process or other point processes that use the average base rate as the only model parameter. This can be very helpful when modeling processes that have the same number of activities, but with different activity occurrence distributions. To demonstrate this ability in Hawkes processes, we show the event occurrence patterns of two simulated Hawkes processes with the same number of activities, but different parameters in Figure 1. We can see that process 1 has more bursty but less regular occurrences compared to process 2, in which less burstiness but a higher regularity is observed. Since both simulated Hawkes processes have the same number of activities in the history, a Poisson model is not able to capture such differences because the base rates of the two processes would be the same in a Poisson process.

For an educational application, there is a natural mapping between student activity events and Hawkes processes. The smaller student activity chunks toward a goal or deadline can

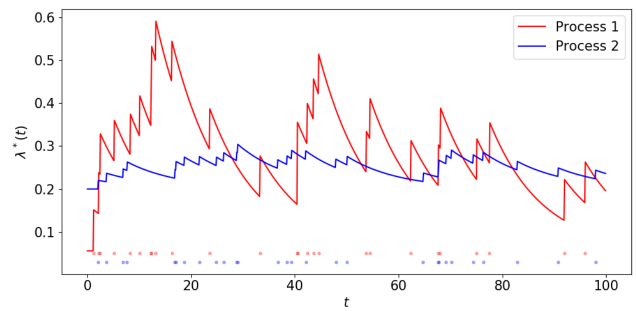

Figure 1: An illustration of two different processes that have the same number of occurrences. The x-axis is the time and the y-axis is the intensity of event occurrences per time unit. Both processes have 29 occurrences but with very different characteristics that will be ignored by Poisson processes.

be examples of an immigrant's offsprings: students break down the big tasks (the whole process) into small sub-tasks (latent clusters). The deadline (external stimuli) of a big task, such as an assignment deadline, can trigger subsequent activities that are associated with the small tasks. These activities arrive closely one after another in a so-called bursty manner (self-excitement) [3]. We demonstrate that Hawkes processes are a good fit to our application by showcasing two examples. First, we show that the module-student pair interactions can not be properly modeled by processes that only model an average base rate, such as Poisson processes. To do this, we conduct a goodness of fit test on the inter-arrival times of module-student pairs in our dataset against the inter-arrival time distribution of a Poisson process, which is exp(1). We use the Kolmogorov-Smirnov test to evaluate the fit's significance. The mean p-value of this test among all module-student pairs in our dataset is $2.77E-6$ with a standard deviation of $6.41E-5$, which shows that module-student pairs do not fit Poisson processes.

Second, we empirically demonstrate the burstiness of module-student interactions. To do this, we show that the Poissonian property of only having a constant base rate is not present in the observed activities of a sample module-student pair from our dataset. Specifically, we use the 1-lag autocorrelation of activity inter-arrival times to conduct our test. The inter-arrival time is defined as the difference between the arrival times of two consecutive activity occurrences. We first simulate a Poisson process with the base rate equal to the average number of activities in our sample activity sequence. Then, we compare the 1-lag autocorrelation in this simulated sequence with the autocorrelation of our sample sequence. Since all inter-arrival times in Poisson processes follow exp(1), we expect the autocorrelation of the simulated Poisson process to be 0 (no correlation). In contrast, we expect to see a non-zero autocorrelation in

[3] It is worth noting that in regular applications of Hawkes processes an activity at time t can trigger later activities at times $\tau > t$. However, in our application, student activities are triggered by the upcoming deadlines in the future. Similarly, earlier chunks of studying sub-tasks at times $\tau < t$ can be offsprings of future studying tasks at time t towards a deadline. As a result, to make the Hawkes process applicable to our problem, we use a reversed activity timeline for our data. This does not affect our model, optimization, or learned parameters.

a bursty self-exciting sequence. Figure. 2 shows the scatter plot of activity inter-arrival times in the original sequence vs. the sequence with lag 1 for each of the two sequences. As we can see, little autocorrelation is spotted in the Poisson process, whereas the pattern of autocorrelation in real data is shown to be not random. Specifically, we can see that most of the lag-1 vs. original inter-arrival times for the sample sequence are scattered around the axes, meaning that dense activities are often followed by long pauses, and vice versa.

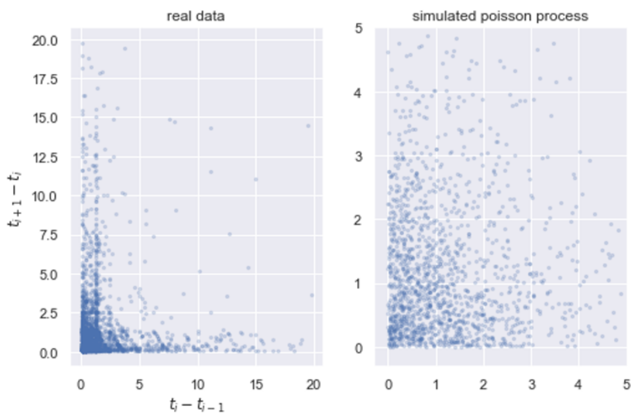

Figure 2: A demonstration of burstiness presented in the interactions of a sample module-student pair: 1-lag autocorrelation scatter plots shows that long pauses are often observed after dense and bursty activities.

It is worth mentioning that our goal is not to directly compare to Poisson models. Rather, we are demonstrating here that we must model the data in a way that captures long-term temporal properties of the processes and their irregularities, rather than static measures such as the count or average number of activities that only provide one facet of the whole picture.

5. METHOD: MULTI-DIMENSIONAL HAWKES PROCESSES TO MODEL ACTIVITY-TYPE RELATIONS

In this section, we introduce the method we use in this study to model student behaviors. More specifically, we illustrate multi-dimensional Hawkes processes and how we apply them to our application. The previous section illustrated how Hawkes process is a good fit for student activities as future activities in module-student pairs could be related to the past activities. In those illustrations, all activities in a module-student pair are considered to be homogeneous or of one single type. In other words, the self-exciting property of the interactions between students and module are assumed to be uniform throughout different kinds of activities, whether it is watching a video lecture, participating in a discussion, or attempting to submit a solution to an assignment. However, in reality, students might exhibit different learning behaviors or use different learning strategies towards different types of learning materials. For example, some students may have more intense and frequent activities when viewing module lectures but less frequent pace when it comes to the discussions. Furthermore, when a student is interacting with two different types of learning materials, different time dependencies may exist between student's interactions with the two. For example, a student may often visit discussion forums very closely after viewing lecture slides, i.e. strong time dependency between lectures and discussions (more specifically, discussion after lectures), but such dependencies may be less obvious for another student.

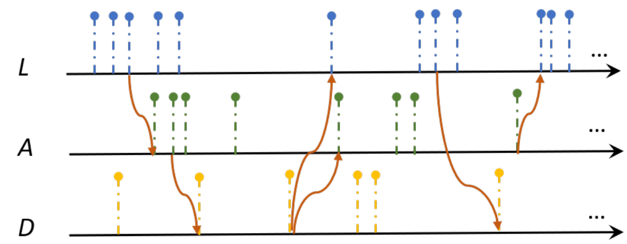

Figure 3: Hawkes processes in module-lecture dimension L, module-assignment dimension A, and module-discussion D, and their mutual excitation. A vertical bar is the representation of an activity occurrence and a red arrow shows the influence of one activity (head) on another (end).

To address this challenge, we model the students' activities on one type of learning material as an individual Hawkes process, and model all such processes simultaneously as multi-dimensional Hawkes Processes. In particular, in the rest of this study, we refer the collection of activities that associate with one type of learning material as a Hawkes process *dimension*. A multi-dimensional Hawkes model not only allows dependency between past and future activities within each dimension (i.e. self-excitation) to be modeled, it is also able to capture the possible dependencies between different types of activities (i.e. excitation between dimensions). For example, scenarios such as submitting the first attempt of an assignment and then starting the second attempt (self-excitation), or, posting a question in the discussion forum after watching a video lecture (excitation between dimensions) can be well described by multi-dimensional Hawkes processes.

In this study, based on the learning material types presented in our dataset, we consider 3 dimensions to analyze students' learning behaviors, namely video lecture dimension L, assignments dimension A and discussions dimension D. To illustrate how multi-dimensional Hawkes processes work in modeling between-dimension excitation, in Figure 3, we show 3 sample Hawkes processes that respectively comes from dimensions L, A, and D. Within each dimension, we use vertical dashed lines to represent the occurrences of activities that take place in that dimension[4]. Activities in one dimension can trigger other activities in another dimensions. This constitutes the influence between different dimensions. We indicate the between-dimension triggers by the red arrows that point from the parent activity to the offspring. For example, the third activity in dimension D in this figure triggers the fifth activity in dimension A as well as the sixth activity in dimension L.

We now formally explain the multi-dimensional Hawkes model and how it can be interpreted according to our application. Suppose that for each module-student pair (m, u), we are

[4]The height of each bar does not represent intensity and does not have any particular meaning in this figure

given a sequence of arrival times for N_{mu} number of activities that are associated with module m and student u. We represent the sequence of each module student pair as in $(m, u) = \{\tau_i\}_{i=1}^{N_{mu}}$, where $\tau_i = (t_i, d_i)$ corresponds to the arrival time of i^{th} activity and the dimension (activity type) d_i to which activity i belongs. For example, suppose student u has 3 total activities in module m. If u submitted an assignment at time 1, then checked a lecture's slides at time 5, and had some discussion posted at time 8, then $(m, u) = \{(t_1 = 1, d_1 = A), (t_2 = 5, d_2 = L), (t_3 = 8, d_3 = D)\}$, with A, L and D representing assignments, video lectures, and discussions respectively. For each dimension $d \in [L, A, D]$ and each module-student pair (m, u), we further use the sequence $T_d(\tau_i) = \{t_i \in \tau_i | \tau_i \in (m, u), d_i = d\}$, to represent the type d learning activities that student u performs in module m as a process. According to the multi-dimensional Hawkes model, we can explain the intensity of $T_d(\tau_j)$ according to the following function:

$$\lambda_d(t) = \mu_d + \sum_{d', t_j < t} \phi_{dd'}(t - t_j), \tag{1}$$

where μ_d describes the average number of activities occurred per unit time that are triggered by exogenous stimuli (the process's base rate in dimension d); and ϕ (the kernel function) represents the function that explains the endogenous stimuli, or the triggering effects from the previous ($t_j < t$) activities in the same dimension or another dimension (d'). In other words, $\phi_{dd'}$ controls the total influence that dimension d exerts on dimension d', as a function of activity interarrival times $(t - t_j)$. Using an exponential kernel function for ϕ, the multi-dimensional Hawkes model can be rewritten as in Equation 2.

$$\lambda_d(t) = \mu_d + \sum_{d'} \alpha_{dd'} \sum_{t_j < t} \beta \exp(-\beta(t - t_j)). \tag{2}$$

The term $\alpha_{dd'}$ and the term $\beta \exp(-\beta(t - t_j))$ can be considered as the decomposition of kernel function $\phi_{dd'}$, which respectively describe the influence weight of dimension d on dimension d' (including α_{dd}, the self-excitation of dimension d itself) and an exponential decay function $g(t) = \beta \exp(-\beta(t))$. Putting together all activity types' parameters, we use a d-dimensional vector $\mu = [\mu_d]$ to represent the base rates of the processes in all dimensions, and a $d \times d$ matrix $\Phi = [\phi_{dd'}]$ to represent the between and within dimension triggering effects. From here, we can write $\Phi = I \circ G$, where we have *influence matrix* $I = [\alpha_{dd'}]$ and *exponential decay kernel* $G = [\sum_{t_j < t}(-\beta \exp(-\beta(t - t_j))]$. Based on that, we can also describe the aggregated influence of dimension d on other dimensions using the following equation:

$$\alpha_d = \frac{1}{|\{d\}|} \sum_{d'} \alpha_{dd'}, \tag{3}$$

which is simply the average influence of dimension d over all dimensions. A summary of all notations used so far are shown in Table 1.

This intensity function $\lambda_d(t)$ has an intuitive meaning: all the future activities in dimension d, apart from those that are triggered by external stimuli, can be triggered by the previous activities that belong to each of the dimensions d' (including d itself) according to the influence weight $\alpha_{dd'}$ (the outer summation). The ones that are triggered by external stimuli take places with rate μ_d. Furthermore, as a past activity becomes distant (larger $t - t_j$), its effect on the occurrence probability of a new event decreases exponentially (i.e. the inner summation). From the branching process point of view, the kernel function $\phi_{dd'}$ is designed in this way so that $\alpha_{dd'}$ is the branching ratio. By computing $\frac{1}{1-\alpha_{dd'}}$, we can obtain the expected number of future activities in dimension d' that are triggered by an immigrant in dimension d. This represents the size of an offspring cluster.

Notation	Description	Formula
L	Dimension module lectures	
A	Dimension module assignments	
D	Dimension module discussions	
(t_j, d_i)	activity j in dimension d_i	
τ_i	arrival time	(t_i, d_i)
(m, u)	module m, student u pair	$\{t_i\}$
$T_d(\tau_i)$	activities in dimension d	$\{t_i \in \tau_i \| d_i = d\}$
μ_d	base rate in dimension d	
$\alpha_{dd'}$	influence of d to d'	
α_d	Influence of dimension d	$\frac{1}{\|\{d\}\|} \sum_{d'} \alpha_{dd'}$
β	decay parameter	
$g(t)$	decay kernel function	$\beta \exp(-\beta(t))$
$\phi_{dd'}(t)$	Hawkes kernal function	$\alpha_{dd'} g(t)$
I	influence matrix	$[\alpha_{dd'}]$
G	decay kernel matrix	$[\sum_{t_j}(-\beta \exp(-\beta(t - t_j))]$
Φ	Hawkes kernel matrix	$I \circ G$
λ_d	intensity in d	Equation 1

Table 1: Notations and their descriptions.

To avoid possible confusions, we also want to clarify that by saying one activity i in dimension d' *triggers* another activity j in dimension d, we mean that the probability of activity j in the result of activity i is higher than j coming from base rate μ_d or triggered by other activities. To see this, one can interpret Equation 1 as follows: in dimension d, a sequence of activities come from the base rate μ_d and each summation leads to a sequence of activities with parameter $\phi_{dd'}$. Then, the probability of j being triggered by i is

$$P(j \text{ child of } i) = \frac{\phi_{dd'}(t_j - t_i)}{\mu_d + \sum_{t_i < t_j} \phi_{dd'}(t_j - t_i)}. \tag{4}$$

Parameter Estimation. A common way to find the best parameters of Hawkes model, given the observed activity arrival times, is to minimize the negative log-likelihood of the data. Particularly, given the sequence $\{(t_i, d_i), ...(t_N, d_N)\}$ till some time T, the log-likelihood of having influence matrix I and base rate vector μ is of the following form:

$$\mathcal{L}(I, \mu) = \sum_{i=1}^{N} log\big(\mu_d + \sum_{t_j < t} \alpha_{dd'} g(t_i - t_j)\big) - T \sum_{d=1}^{n} \mu_d - \sum_{d} \sum_{d'} \alpha_{dd'} \int_0^{T - t_j} \phi(T - t_j) dt_j. \tag{5}$$

In order to find Hawkes parameters that models each module-student pair, we adopted algorithm ADM4 [45], which made use of a mix of Lasso and nuclear regularization on top of the negative log-likelihood. Specifically, Accelerated Projected Gradient Descend method was used to meet the non-negative constraints on I and μ as Hawkes parameters only have realistic meanings when the parameters are non-negative[5]. When it comes to the selection of global parameter β, for

[5]We made our implementation available at https://github.com/ssahebi/EDM2020-Hawkes.

each module-student pair, we use grid search with cross validation on the interval $[0, 10]$ with step size 1.

6. EXPERIMENTS

6.1 Testing the Goodness of Fit

To test the goodness-of-fit of the model, in Table 2, we compare the RMSE of the intensity for all dimensions, computed based on the observed module-assignment pairs, for multi-dimensional Hawkes model (i.e. Equation 1), single-dimensional Hawkes model and a Poisson model. Specifically, for single-dimensional Hawkes, we treat all activities as in one dimension, and estimate the intensity of each dimension using the uni-variate parameters α, β, and μ. For the Poisson model, in each dimension, we use the average activity arrival rate as the base rate, and compute the RMSE for each dimension respectively. As we can see in

	L	A	D
Hawkes (Multi)	0.56	2.34	1.37
Hawkes (Single)	0.71	2.57	1.95
Poisson	3.22	6.91	3.73

Table 2: The goodness of fit to true data for each model in terms of intensity RMSE.

Table 2, Poisson has the worst fit in all dimensions, possibly caused by the non-Poissonian nature we showed from the real data. Single-dimensional Hawkes has comparable but slightly worse performance. One possible reason is that there might exist differences in terms of base rate and burstiness between dimensions and by modeling all types of learning materials as one activity type, the model can only capture the average trend in all dimensions. To visualize how the multi-dimensional Hawkes processes fit the real data, we also present in Figure 4 the estimated intensity (blue) and true intensity (black) of a sample module-student pair. As we can see in this figure, the model mostly has a good fit to the real data Only at some time points, it underestimates the expected number of activities that are about to happen.

6.2 Model Parameter Analysis: Trends and Differences Between Dimensions

In this section, we analyze the estimated Hawkes parameters within and between different dimensions, to show their general trends and differences across different dimensions.

We start this part with a correlation analysis of all Hawkes Parameters, to show the general trends and possible differences between dimensions. Particularly, we calculate the Spearman rank correlation coefficients between the parameters that are learned for all module-student pairs as is shown in Figure 5. Recall that parameters $\alpha_{dd'}$, μ_d, β and α_d respectively is the between-dimension (or within if $d = d'$ excitation, base rate, decay rate (Equation 2) and aggregated influence of dimension d (Equation 3) for $d \in [L, A, D]$.

We can see that self-excitation within dimensions (i.e.α_{dd}) are generally negatively correlated with base rates μ_d of the same dimensions and decays β. This means that as the external stimuli leads to more and more expected arrivals, i.e. when regular activities come from the base rate, the effect of each previous activity on the future ones tends to decrease, i.e. self-excitation gets weaker. In other words, in

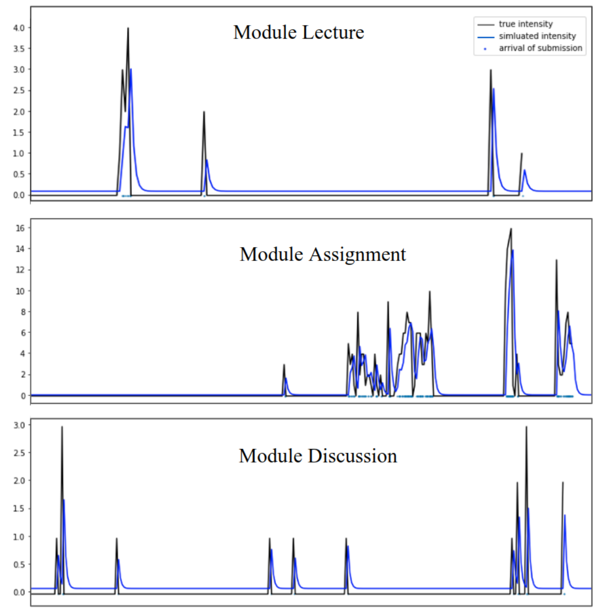

Figure 4: Estimated and true intensity of a sample module-student pair, modeled by multi-dimensional Hawkes.

sequences with higher regularity, less burstiness is observed. Also, it means that activities that have a slower decay rate usually arrive in a more bursty manner. Mapping to our application of students interacting with learning modules, by using the branching process point of view, higher α suggests higher expected number of activities in a latent cluster as sub tasks. On the other hand, the negative correlation also means a lower base rate and lower number of immigrants. In other words, the number of such latent clusters are also fewer. One possible interpretation is that in each dimension, students divide their big learning task into sub tasks and work for each individual sub task in a relatively bursty manner. This also suggests that students barely have behaviors that are both highly intense and highly frequent (i.e. large μ_d and α_{dd}). Similarly, both highly sparse and highly mediated activities (i.e. small μ_d and α_{dd}) are rarely observed neither, as the correlation between μ_d and α_{dd} is positive for all $d \in [L, A, D]$.

Comparing the parameter correlations across different activity types, we can see that dimensions L and D, have high within and between-dimension influence correlations. For example, the correlation between α_{LL} and α_{LA} is 0.97 in dimension L and correlation between α_{DL} and α_{DA} is 0.96. This implies that the influence of discussion and video lecture activities on other dimensions are almost consistent. For example, if the influence of video lectures on assignments is high, it is likely that the influence of video lectures on discussions is high too. Similarly, if the pattern of interacting with video lectures in a module is bursty (high α_{LL}), it is likely that other activity types triggered by video lectures are also bursty. However, the influence of assignment activities on discussions and video lectures are not significantly associated with assignment activity's self-excitement. That could mean that after a student has a bursty set of assignment activities in a module, the student is less likely to have a bursty video lecture activities. Similarly, the influence of assignment activities on discussions has a low correlation with the influence of assignment activities on video lectures. For instance, if a student starts an assignment intensively very closely after some intensive watching of video lectures, the student is less likely to have high intense discussion activities after assignments. We can also see that assignment-triggered activities' burstinesses are less correlated with the base rates (i.e. α_{AD} vs. μ_d). This means that the frequency of activities that come from external stimuli does not affect the influence of assignment activities on consequent activities in other dimensions. Taken altogether, it is interesting to see that activities that are associated with assignments tend to have different exciting patterns compared to video lectures and discussions. This can show the influence of deadlines, as the assignments are the only activity type that have deadlines and are going to reflect student grades in this dataset.

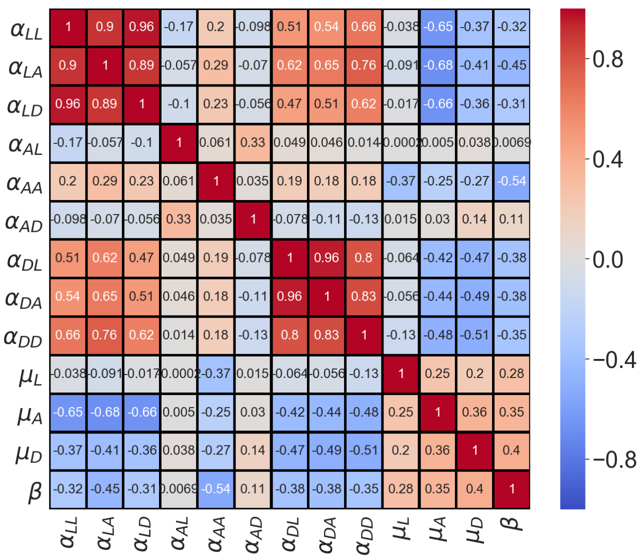

Figure 5: Spearman Rank Correlation Coefficients between Hawkes Paramters.

6.3 Student Behaviors Characterized by Model Parameters

In the previous part, we were interested in showing the correlation between Hawkes parameters that represented within and the between-dimension relationships. In this part of the analysis, we focus on the different behaviors that are observed according to the learned parameters for each module-student pair. Additionally, we are interested to see if these learned Hawkes parameters are proper representatives for student procrastination. To do so, we first define a measure that can represent student procrastination in the absence of self-reported data. In the following, we go over some important assumptions, definitions and time measures that we use for procrastination.

Defining Delay as a Procrastination Measure. We assume that each student works on one module at a time, meaning that they do not work on several modules at the same time. Furthermore, we assume that submitting the last attempt of the module's assignment marks the end of studying this module. According to these assumptions, we define module i's end time for student j, t^e_{ij} as the time stamp when the last module assignment was submitted. We then define the start time t^s_{ij} as the earlier time stamp between $t^e_{i-1,j}$ and the available time for module i. In other words, when student j finishes learning module $i-1$, if module i has already been made available, then their end time on module $i-1$ is defined as the start time for module i. Otherwise, the start time is going to be the time when module i becomes available or is published online. In each dimension d, we use t^d_{ij} to denote action time, which is defined as the time that the first activity in dimension d takes place between start time t^s_{ij} and end time t^e_{ij}.

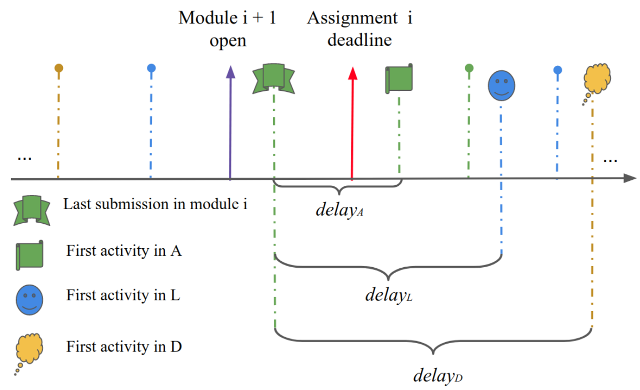

Figure 6: Illustration of delay measures. As in Figure 3, we use blue, green and yellow dashed lines to represents activities from dimension L, A and D respectively.

Having the module start time and student action time in dimension d, we can calculate how late a student started working on activity of type d in the module using $t^d_{ij} - t^s_{ij}$. To factor in the duration differences between different modules, we normalize this value by the module duration. Eventually, we define the following measure to quantify student j's normalized delay in dimension d that is associated with module i:

$$delay_d = \frac{t^d_{ij} - t^s_{ij}}{t^e_{ij} - t^s_{ij}} \quad (6)$$

One of the motivations to define the delay according to start time t^s_{ij} is that, sometimes module $i+1$ is available before the assignment deadline in module i. By this time, student j might still be working on module i. So, it would be unfair to count the time after module $i+1$ is available and before student j's assignment submission as the procrastinating time for student j on starting module $i+1$. On the other hand, if student j finishes the assignment in module i earlier than the deadline of the assignment, this extra time they earned from the early submission can be used toward the next available module. If the student does not use this time, it will be considered as a cramming behavior toward the next module $i+1$. An illustration of these definitions is presented in Figure 6.

Observing Two Behavior Groups. We now focus on the distribution of the learned Hawkes parameters to see if we can observe any behavioral patterns across different student-module pairs. Specifically, in Figure 7 we present the distribution of the learned α_{LL}, α_{AA}, and α_{DD}. We can clearly observe two spikes in the density distribution of influence parameters, more prominently in α_{LL} and α_{AA}. Combining this observation with the correlation analysis in previous section, we can see that there are two types of module-student interactions: the ones with higher frequency and lower burstiness versus the ones with lower frequency and higher burstiness. To statistically show the difference between these two types of interactions, we first cluster student-modules according to their α_{LL} and α_{AA}, into two groups using the K-means clustering algorithm. Then, we test to see if the learned Hawkes parameters, i.e. excitation parameters $\alpha_{dd'}$, base rate μ_d, decay β and aggregated influence α_d for $d \in [L, A, D]$, are statistically different across the two groups. Particularly, we conduct the Kruskal-Wallis test on each learned parameter between the two clusters. The average values of the parameters for each of the two clusters are shown in Table 3. Since the p-values for all tests are smaller than 0.0001, we do not show them in the table. These small p-values suggest that the differences between clusters are statistically significant for all parameters between the two types. This indicate that the differences between the two types are meaningful.

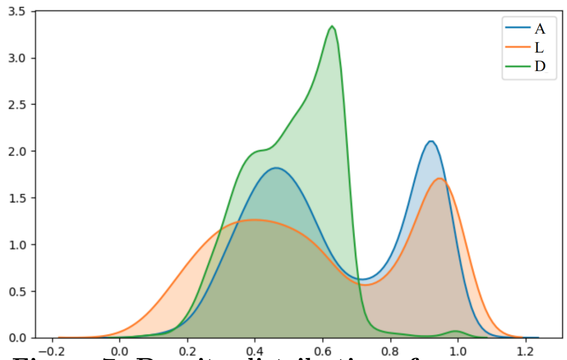

Figure 7: Density distribution of α_M, α_A, α_D.

Examining both groups more closely, we can see that, the aggregated influence of dimension A, i.e. α_A, is the highest among all 3 for both type 1 and type 2 groups. With that being said, this influence majorly comes from the self-excitement in dimension A, i.e. α_{AA}. μ_A is also the highest among all 3 dimensions. Combining these observations, we can see that assignment-related activities arrives more frequently and are highly influential in triggering consequent activities, especially the assignment-related ones. Also, we can see that on average type 2 group has a much smaller base rate for video lectures (μ_L) and discussions (μ_D), meaning less density and regularity in those activities compared to type 1. However, In assignment-related activities, the base rate (μ_A) as well as aggregated influence in dimension A (α_A), are higher in type 2 group, which suggests an overall denser and more intense assignment-related activities arrivals comparing to type 1 group.

Now if we look at the differences between two groups in terms of between-dimension relationships, we can see that α_{AL}, i.e. the triggering effect of assignment to consequent video lecture activities (and similarly, α_{AD}: the assignment-triggered discussion activities) is much lower in type 1 group compared to type 2 group. This difference is also notable in other between-dimension αs. For example, the influence of assignments on video lectures (α_{AL}) is way less than the influence of video lectures on discussions (α_{LD}) in type 2 group, while this difference is less in the type 1 group. This suggests that the interaction patterns with assignments in type 2 group are almost inconsistent with other dimensions. We note that, although the type 1 and type 2 clusters are created according to α_{LL} and α_{AA} parameters only, we see significant differences in all other parameters of the two groups.

Delay in the Discovered Groups. Here, we aim to understand if the two behavior types that we discovered in the previous part are associated with measures of procrastination. Particularly, we evaluate the differences observed in the delay measures defined in Equation 6 for the two clusters. The results are presented in Table 4. Again, all p-values are smaller than 0.0001. A major observation is that type 1 and type 2 have very different delays in all dimensions. Specifically, the delay of each dimension in type 1 group is much less than the corresponding delays in type 2 group. As a result, we can call type 2 group as the delay group and type 1 group as the non-delay group. Given that the type 1 and type 2 behavioral clusters are formed based on Hawkes model parameters only, this important observation demonstrates that the learned Hawkes parameters can clearly represent delay as a procrastination measure. Also, we can see that in delay (type 2) group, on average, students start the first discussion way after the first assignment activity takes place ($delay_D > delay_A$). However, in the non-delay (type 1) group, on average the first assignment activity happens after some discussion ($delay_D < delay_A$). We can see that in both groups, the video lecture activities come before discussions or assignments.

Combined with our observations from the previous analysis, we see that not only the delay group start the first activity in each dimension much later than the other group, they also have a much less base rate μ_L and μ_D. Consequently, we can see that the delay group (type 2) has less frequent but more bursty discussion and lecture-related activities, while the non-delay group activities arrive in a less bursty but more frequent manner in these two dimensions. On the other hand, assignment activities are denser and more intense for the delay group. This combined observation shows that the Hawkes parameters can represent more information about student procrastination, compared to the delay measure alone.

6.4 Student Grades Associated with Model Parameters

In the previous section, we concluded that the learned Hawkes model parameters not only represent delays, but also can capture more procrastination-related behaviors. In the rest of this section, we are interested in exploring if the additional trends captured by the Hawkes model can be more meaningful in association with student grades, compared to the delay parameter. In particular, we are interested in the association between delay and student grades from the Hawkes processes point of view.

Recall that $delay_d$ defined in Equation 6 measures the normalized delay of the first activity in dimension d of the

	d	α_{dL}	α_{dA}	α_{dD}	α_d	μ_d	β
Type 1	L	0.558±0.149	0.263±0.178	0.289±0.185	0.381±0.524	0.0003±0.0006	0.663±0.692
	A	0.107±0.272	0.820±0.125	0.101±0.264	0.462±0.459	0.0003±0.002	
	D	0.322±0.393	0.305±0.393	0.790±0.151	0.394±0.325	5.52E-5±1.8E-4	
Type 2	L	0.874±0.108	0.823±0.135	0.816±0.134	0.795±0.582	4.82E-5±9.86E-5	0.425±0.249
	A	0.019±0.124	0.864±0.061	0.018±0.124	0.799±0.582	0.0004±0.004	
	D	0.699±0.429	0.696±0.430	0.936±0.092	0.590±0.238	1.19E-5±5.92E-5	

Table 3: Statistics of Hakwes parameters $\alpha_{dd'}$, μ_d, β and α_d for $d \in [L, A, D]$ in the two clusters.

	$delay_L$	$delay_A$	$delay_D$
Type 1	0.08±0.228	0.575±0.411	0.338± 0.385
Type 2	0.108±0.274	0.722±0.360	0.819±0.337

Table 4: Statistics of delay measures $delay_L$, $delay_A$ and $delay_D$ in two clusters identified by Hawkes parameters.

d	L	A	D	avg.
$delay_d^H$	-0.339***	-0.125*	-0.329***	-0.264**
$delay_d$	-0.240**	-0.070.	-0.114*	-0.141*

Table 5: Spearman's correlation with respect to assignment score for each delay measure. Significance level is denoted as follows: p<0.001* p<0.01 ** p<0.05* p<0.1.**

student-module pair. Here, we define a new delay measure based on both learned parameters of the Hawkes process and $delay_d$. We then study if this newly defined delay measure performs better in association with student grades, compared to $delay_d$. Specifically, after showing the between-dimension excitation interrelationships, it is reasonable to assume that these interrelationships are important in the activity delays as well. For example, knowing that assignment-related activities can trigger followup activities in all 3 dimensions, delaying the assignment-related activities also potentially causes consequent delays in other dimensions. Motivated by this, we propose $delay_d^H$ by combining $delay_d$ and between-dimension Hawkes parameters as follows:

$$delay_d^H = delay_d + \frac{1}{\sum_{d' \neq d} \frac{1}{1-\alpha_{dd'}}} \tag{7}$$

As we mentioned in Section 5, $\frac{1}{1-\alpha_{dd'}}$ can be seen as the statistically expected number of activities in a latent cluster that are triggered by an immigrant. The second term in Equation 7 basically quantifies the potential loss per time unit in terms of triggering other dimensions' activities by delaying in dimension d. Taken altogether, $delay_d^H$ describes the total delays in all 3 dimension that are associated with delay in dimension d.

To see if $delay_d^H$ provides more grade-related information compared to $delay_d$, we look at the Spearman's correlation between these two measures and students' assignment grades. The result of this correlation analysis is presented in Table 5. Our first observation is that the correlations between both delay measures with student grade are negative. However, this correlation is not as significant for $delay_d$, compared to $delay_d^H$. This is specially stronger in the assignment dimension. The reason for this can be two-fold: (1) comparing to $delay_d$, $delay_d^H$ not only captures how late the action was taken in each dimension, it also provides some insights on the student behavior trends throughout their learning process, and (2) as $delay_d^H$ describes the time-dependencies between dimensions, it is more powerful in explaining student activities in all dimensions as a whole, compared to the point estimate summaries of procrastination. Particularly, one may overlook the importance of delaying the discussion-related activities on assignment grades when considering the $delay_D$ measure only. However, a stronger correlation between $delay_D^H$ and grades suggests that early start of the discussion-related activities is almost equally important as starting the video lectures early, probably because of the triggering effect of dimension D and the potential loss that its delay causes to all 3 dimensions.

7. CONCLUSION

In this work, we proposed to use the multi-dimensional Hawkes processes to model procrastination in student learning behavior. We showed that multi-dimensional Hawkes processes have a better fit to student activity counts in comparison with their single-dimensional version and the Poisson processes. By analyzing the correlations between the learned parameters in the Hawkes processes, we concluded that more bursty student sequences have less regular activities in them, the burstiness of video lecture and discussion-related activities vary similar to each other, and the deadlines highly affect the arrival times of assignment-related activities. We showed that Hawkes parameters can reveal two types of behaviors in the data that are associated with different delays - the delay group tends to have high within and between-dimension excitation but low base rate, and the non-delay group have a high base rate and a lower excitation in all dimensions. According to the branching processes point of view, we gave a realistic interpretation on these types of behaviors: non-delay group divide big tasks into many subtasks (high base rate) which leads to more frequent and less dense activities throughout the learning process. On the other hand, delay group tend to intensively work in one dimension for a shorter period of time, followed by long pauses (high excitation but low base rate). We also showed that the Hawkes model parameters represent richer information compared to the delay measure alone by defining a new Hawkes-based delay measure and associating it with student grades. Our experiments demonstrated that the between-dimension dependencies in the multi-dimensional Hawkes model better explain student grades.

This study is limited in the number and variety of the datasets that we have experimented on. In the future, we plan to explore more datasets from various disciplines and platforms. Another limitation is the single measure that we use to evaluate procrastination ($delay_d$). As a followup to this study, we aim to define and use more procrastination indicators, including the self-reported procrastination measures.

8. ACKNOWLEDGEMENT

This paper is based upon work supported by the National Science Foundation under Grant Number 1917949.

9. REFERENCES

[1] J. W. Alstete and N. J. Beutell. Performance indicators in online distance learning courses: a study of management education. *Quality Assurance in Education*, pages 6–14, 2004.

[2] C. J. Asarta and J. R. Schmidt. Access patterns of online materials in a blended course. *Decision Sciences Journal of Innovative Education*, 11(1):107–123, 2013.

[3] R. Azevedo and R. Feyzi-Behnagh. Dysregulated learning with advanced learning technologies. volume 7, pages 9–18. Italian e-Learning Association, May 2011.

[4] C. Backhage, C. Ojeda, and R. Sifa. Circadian cycles and work under pressure: A stochastic process model for e-learning population dynamics. In *Data Science–Analytics and Applications*, pages 13–18. Springer, 2017.

[5] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[6] R. Baker, B. Evans, and T. Dee. A Randomized Experiment Testing the Efficacy of a Scheduling Nudge in a Massive Open Online Course (MOOC). *AERA Open*, 2(4):233285841667400, Oct. 2016.

[7] P. Bao, H.-W. Shen, X. Jin, and X.-Q. Cheng. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes. In *Proceedings of the 24th International Conference on World Wide Web*, pages 9–10. ACM, 2015.

[8] K. Börner, O. Scrivner, M. Gallant, S. Ma, X. Liu, K. Chewning, L. Wu, and J. A. Evans. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences*, 115(50):12630–12637, 2018.

[9] R. Cai, X. Bai, Z. Wang, Y. Shi, P. Sondhi, and H. Wang. Modeling sequential online interactive behaviors with temporal point process. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 873–882, 2018.

[10] R. Cerezo, M. Esteban, M. Sánchez-Santillán, and J. C. Núñez. Procrastinating Behavior in Computer-Based Learning Environments to Predict Performance: A Case Study in Moodle. *Frontiers in Psychology*, 8, Aug. 2017.

[11] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, 96:42–54, May 2016.

[12] J. N. Choi and S. V. Moran. Why not procrastinate? development and validation of a new active procrastination scale. *The Journal of social psychology*, 149(2):195–212, 2009.

[13] R. W. Crues, N. Bosch, M. Perry, L. Angrave, N. Shaik, and S. Bhat. Refocusing the lens on engagement in moocs. In *Proceedings of the fifth annual ACM conference on learning at scale*, pages 1–10, 2018.

[14] N. Du, Y. Wang, N. He, J. Sun, and L. Song. Time-sensitive recommendation from recurrent user activities. In *Advances in Neural Information Processing Systems*, pages 3492–3500, 2015.

[15] T. Dvorak and M. Jia. Do the Timeliness, Regularity, and Intensity of Online Work Habits Predict Academic Performance? *Journal of Learning Analytics*, 3(3):318–330, 2016.

[16] G. C. Elvers, D. J. Polzella, and K. Graetz. Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology*, 30(2):159–162, 2003.

[17] B. J. Evans, R. B. Baker, and T. S. Dee. Persistence patterns in massive open online courses (moocs). *The Journal of Higher Education*, 87(2):206–242, 2016.

[18] B. Gelman, M. Revelle, C. Domeniconi, A. Johri, and K. Veeramachaneni. Acting the same differently: A cross-course comparison of user behavior in moocs. *International Educational Data Mining Society*, 2016.

[19] P. F. Halpin and P. De Boeck. Modelling dyadic interaction with hawkes processes. *Psychometrika*, 78(4):793–814, 2013.

[20] P. F. Halpin, A. A. von Davier, J. Hao, and L. Liu. Measuring student engagement during collaboration. *Journal of Educational Measurement*, 54(1):70–84, 2017.

[21] H. Jing and A. J. Smola. Neural survival recommender. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 515–524, 2017.

[22] A. M. Kazerouni, S. H. Edwards, T. S. Hall, and C. A. Shaffer. DevEventTracker: Tracking Development Events to Assess Incremental Development and Procrastination. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE '17*, pages 104–109, Bologna, Italy, 2017. ACM Press.

[23] K. R. Kim and E. H. Seo. The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, 82:26–33, 2015.

[24] R. F. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 170–179, 2013.

[25] A. S. Lan, C. G. Brinton, T.-Y. Yang, and M. Chiang. Behavior-based latent variable model for learner engagement. *International Educational Data Mining Society*, pages 64–71, 2017.

[26] A. S. Lan, J. C. Spencer, Z. Chen, C. G. Brinton, and M. Chiang. Personalized thread recommendation for mooc discussion forums. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 725–740. Springer, 2018.

[27] C. H. Lay. At last, my research article on procrastination. *Journal of research in personality*, 20(4):474–495, 1986.

[28] N. Michinov, S. Brunot, O. Le Bohec, J. Juhel, and M. Delaval. Procrastination, participation, and performance in online learning environments. *Computers & Education*, 56(1):243–252, 2011.

[29] S. Mishra, M.-A. Rizoiu, and L. Xie. Feature driven and point process approaches for popularity

prediction. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1069–1078. ACM, 2016.

[30] S. M. Moon and A. J. Illingworth. Exploring the dynamic nature of procrastination: A latent growth curve analysis of academic procrastination. *Personality and Individual Differences*, 38(2):297–309, 2005.

[31] C. Network. Canvas network courses, activities, and users (4/2014-9/2015) restricted dataset. *Harvard Dataverse*, 1, 2016.

[32] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[33] J. Park, R. Yu, F. Rodriguez, R. Baker, P. Smyth, and M. Warschauer. Understanding student procrastination via mixture models. *International Educational Data Mining Society*, 2018.

[34] L. W. Perna, A. Ruby, R. F. Boruch, N. Wang, J. Scull, S. Ahmad, and C. Evans. Moving through moocs: Understanding the progression of users in massive open online courses. *Educational Researcher*, 43(9):421–432, 2014.

[35] T. A. Pychyl, J. M. Lee, R. Thibodeau, and A. Blunt. Five days of emotion: An experience sampling study of undergraduate student procrastination. *Journal of social Behavior and personality*, 15(5):239, 2000.

[36] M.-A. Rizoiu, L. Xie, S. Sanner, M. Cebrian, H. Yu, and P. Van Hentenryck. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, pages 735–744. International World Wide Web Conferences Steering Committee, 2017.

[37] A. Rotenstein, H. Z. Davis, and L. Tatum. Early birds versus just-in-timers: the effect of procrastination on academic performance of accounting students. *Journal of Accounting Education*, 27(4):223–232, 2009.

[38] J. Shang and M. Sun. Geometric hawkes processes with graph convolutional recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4878–4885, 2019.

[39] K. Sharma, P. Jermann, and P. Dillenbourg. Identifying styles and paths toward success in moocs. *International Educational Data Mining Society*, 2015.

[40] P. Steel. The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin*, 133(1):65, 2007.

[41] B. W. Tuckman. Relations of academic procrastination, rationalizations, and performance in a web course with deadlines. *Psychological reports*, 96(3_suppl):1015–1021, 2005.

[42] A. A. von Davier and P. F. Halpin. Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series*, 2013(2):i–36, 2013.

[43] W. Xiao, X. Xu, K. Liang, J. Mao, and J. Wang. Job recommendation with hawkes process: an effective solution for recsys challenge 2016. In *Proceedings of the recommender systems challenge*, pages 1–4, 2016.

[44] J. Yoo and J. Kim. Can Online Discussion Participation Predict Group Project Performance? Investigating the Roles of Linguistic Features and Participation Patterns. *International Journal of Artificial Intelligence in Education*, 24(1):8–32, Jan. 2014.

[45] K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional hawkes processes. In *International Conference on Machine Learning*, pages 1301–1309, 2013.

[46] M. Zhu, Y. Bergner, Y. Zhang, R. Baker, Y. Wang, and L. Paquette. Longitudinal engagement, performance, and social connectivity: a mooc case study using exponential random graph models. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 223–230, 2016.

[47] B. J. Zimmerman. Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1):166–183, 2008.