Pathways of Students’ Progress through an On-demand Online Curriculum

Dr. Jim Morgan P.E., Charles Sturt University
Jim Morgan is the father of two daughters and the spouse of an engineer. Before joining Charles Sturt University as Professor of Engineering and Inaugural Course Director in 2015, he was on the faculty in civil engineering at Texas A&M for over 30 years. Jim has been active in the freshman engineering program at A&M for nearly 20 years; was an active participant in the NSF Foundation Coalition from 1993 to 2003; also has received funding for his engineering education research from the Department of Education FIPSE program and from the National Science Foundation CCLI program. He is active in the American Society for Engineering Education, is past chair of the Freshman Programs Division, and has served on the FIE steering committee. In addition to his teaching in engineering, Jim served several years as Co-Director of the Eisenhower Leadership Development Program in the Center for Public Leadership at the George Bush School of Government and Public Service; and also served as director of Aggie STEM with funding from the Texas Education Agency and the Texas Higher Education Coordinating Board.

Prof. Euan Lindsay, Charles Sturt University
Professor Euan Lindsay is a Mechatronic engineer, a discipline that integrates computers, electronics and physical hardware. Prof Lindsay’s background is in Remote laboratories, investigating whether remote and simulated access alternatives to the traditional in-person laboratory experience can provide the high quality learning outcomes for students.

Prof Lindsay’s work in Remote and Virtual laboratory classes has shown that there are significant differences not only in students’ learning outcomes but also in their perceptions of these outcomes, when they are exposed to the different access modes. These differences have powerful implications for the design of remote and virtual laboratory classes in the future, and also provide an opportunity to match alternative access modes to the intended learning outcomes that they enhance.

Prof Lindsay is the Foundation Professor of Engineering at Charles Sturt University. His research interests centre largely around online learning – the use of remote and virtual laboratories, MOOCs and other methods for making learning asynchronous, and data analytics for promoting student learning.

Prof Lindsay was the 2010 President of the Australasian Association for Engineering Education. He is a Fellow of Engineers Australia, and a Fellow of the UK Higher Education Academy. Prof Lindsay was the recipient of a 2007 Carrick Award for Australian University Teaching. In 2005 he was named as one of the 30 Most Inspirational Young Engineers in Australia.

Dr. Colm Howlin, Realizeit
Dr. Colm Howlin is the principal researcher at Realizeit and leads the research and analytics team. He has been with the company since it was founded over 10 years ago. He is responsible for the development of the Adaptive Learning Engine within Realizeit and the Learning and Academic Analytics derived from learner data. Colm has a background in applied mathematics, earning his B.Sc. and Ph.D. in applied mathematics from the University of Limerick. He was a Research Fellow at Loughborough University in the UK. Colm has over 10 years’ experience working on research, educational data, analytics, and statistical analysis, including spending time as a consultant statistician before joining Realizeit.

Dr. Maartje E. D. Van den Bogaard, Delft University of Technology
Dr Van den Bogaard is director of studies and assistant professor of Science Education and Communication at Delft University of Technology, the Netherlands.
Pathways of students’ progress through an on-demand online curriculum

Abstract
Charles Sturt University makes its underpinning technical curriculum available to its students using an on-demand online system they call their Topic Tree. The tree is a directed acyclic graph where nodes represent topics to be learned, and edges represent the prerequisite relationships that exist between the topics. Branches on the topic tree represent concentrations in an area of knowledge, sub-branches (water quality, fluid mechanics, etc.) represent distinct subsets of knowledge - specialty. Delivery of the technical content is in three-hour modules, and students are free to choose the order in which they engage with these topics.

Previous work has identified that students engage with the on-demand curriculum much as they engage with on-demand entertainment platforms such as Netflix, completing long sequences of topics with short periods between them – the traditional “binge” model of consumption.

This paper presents a more fine-grained analysis of students’ pathways through the topic tree, focusing on the distance between successive topics completed by the students. Students’ progress is characterized by a three-dimensional framework – time, distance, and purpose.

In general, pathways through the tree fall into one of four patterns:
- Forward movement along a branch of the tree,
- Movement backward along a branch of the tree,
- Repeating the same topic,
- Switching to a different branch of the tree (backward distance to the junction of the branches combined with a forward distance along the new branch)

Different students engage with the topic tree using different combinations of these pathways, distance absolute distance traveled through the topics, and different time gaps between activities on the topics. This paper will identify the different combinations that can be found in the student log data.

1. Introduction
Charles Sturt University (CSU) Engineering is a new programme established from scratch in 2016 by a university that had not previously taught engineering. This was taken as an opportunity to build an all-new programme structure and philosophy [1]. Students at CSU Engineering complete a sequence of semester-long Project-Based Learning (PBL) style challenges across the first three semesters; after this point, they commence industry-based work placements.

The delivery of the underlying technical curriculum is through the Realizeit platform [2] and is based on a philosophy of self-directed learning. Students have freedom in deciding how, when and, to a large extent, which elements of the curriculum they engage within the online environment. This freedom, along with the PBL-style challenges, is enabled by the structure of the technical curriculum which is broken down into fine-grained learning activities called ‘topics.’ These topics are arranged into a tree structure where the recommended learning order is made explicit [3], see Figure 1. Each topic has its own learning outcomes, learning resources and
assessment. Some are automatically assessed online; others require submissions that are marked by faculty. A topic is intended to take a typical student around three hours to complete.

Figure 1: The CSU Engineering topic tree. The tree contains 689 distinct topics. Student progress from left (the root of the tree) to right (more advanced topics). Colours represent sub-branches/specialties.

This structuring of the curriculum and learning environment allows students to engage with theory at the point it is required in their work, rather than in a synchronous syllabus driven manner. This ensures content is relevant at the point at which it is learned. However, an approach such as this does introduce risks surrounding how students manage their progression through the tree. The work presented in this paper is in part motivated by an interest in students’ behaviour in order to better understand what constitutes ‘at risk’ in this unique environment.

Previous work identified that students engage with the on-demand curriculum much as they engage with on-demand entertainment platforms such as Netflix, completing long sequences of topics with short periods between them – the traditional “binge” model of consumption.
This paper seeks to extend this work to a more fine-grained analysis of students’ pathways through the topic tree, focusing on the distance between successive topics completed by the students. Students’ progress is characterized by a three-dimensional framework – time, distance, and purpose. An exploration of these will form the bulk of this paper.

1.1. Topic Acquisition and Progression

Before students are eligible to go on industry placement in their fourth semester, they are required to complete a minimum of 240 topics from the tree. While 80 of these topics are compulsory, selected after a process of engaging with industry partners [1], the students are free to choose the other 160 from throughout the tree.

The topic tree is a directed acyclic graph where nodes represent topics to be learned, and edges represent the prerequisite relationships that exist between the topics. Branches on the tree comprise of sequences of topics which are strongly related and scaffolded on top of each other. Although there is no required order in which students should complete topics, the structure of the tree provides clear guidance as to the recommended learning path, with sequential topics building upon the learning of the prior topics on their branch.

A topic completion is when a student completes the assessment for that topic, either by working through the automated online assessment in topics where available, or by uploading a submission for manual marking that is later assessed as complete. Uploaded submissions that are assessed as not having met the learning outcomes do not count as completions.

The Topic Tree operates on a mastery learning paradigm, as such they must, through assessment and submission of project work, meet the given criteria for mastery for the platform and instructor to consider a topic complete. Therefore, it is not uncommon for students to require multiple engagements with the learning material, depending on their level of competence. Unlike a traditional mastery learning paradigm, where students can make as many attempts as necessary to reach the required threshold of mastery, students are restricted to at most three assessment submissions for each topic. They have unlimited learning access; but fourth and subsequent assessment accesses do not count for credit.

As the learning is self-directed, progress and the rate at which it occurs is the responsibility of the individual student. There are no immediate consequences for an inadequate rate of progression, and while some measures are taken to signal adequate progression to encourage students to stay on track [4], some students still lag. Lindsay & Morgan [3] observed that this leads to the emergence of two sub-cohorts – one who was up to date, and one that was well behind and unlikely to complete the course.

The challenge for the teachers at CSU Engineering is to develop metrics of behavior that support understanding effective study behavior in this unique curriculum. This paper presents important steps in that direction.

1.2. Examples of Topic Progression

The sample topic tree shown in Figure 2 can be used to illustrate a range of student behaviours.
Students can undertake multiple learning activities within the same topic (eg A->A). A student who has successfully completed a topic may move forward in the tree, either to the next topic (eg A->D), or if they are confident they can skip ahead (eg A->F). The formal sequence is not enforced by the Topic Tree, although anecdotal (unpublished) data shows that skipping prerequisites for a topic appears to lead to longer completion times for that topic.

For a student who is unsuccessful in learning the material, it is common to retreat along the branch, to the immediate previous topic (eg A->T₁), or perhaps to a prior topic (eg A->T₀) if they feel deeper revision is required or if skipping has occurred.

Alternatively, a student may change branches, either to a nearby branch (eg C->D) or a distant branch (eg C->G). This can be triggered by completion of a whole branch, a partial branch that has met a student’s needs, abandonment of a branch that is seen as too hard, or simply a desire to see other parts of the tree. Of these triggers, only the completion of a whole branch can be explicitly observed in this data. Additional metrics, such as scores on questions, would be required to infer the others.

2. Methodology
   2.1. Research Questions
   In our previous study [5], we considered how students move through the topic tree using total distance traveled on the tree and time gap between activities. In this work, we extend the analysis to consider not only the distance traveled, but also the direction (backward, forward, new branch) through the tree and importantly the purpose (learn/revise) of the activity. These additional dimensions allow a more detailed and revealing view of student behaviors and how they engage with the topic tree.

2.2. Dataset
   For this study, we used learning logs from the first three student cohorts of the programme, cohorts 2016, 2017 and 2018. Cohort 2016 and 2017 contain 27 and 28 students respectively and have started their industry placements. Cohort 2018 contains 19 students who are in the initial phase of their programme. As the latest cohort has only started interacting with the learning platform, they contribute the smallest amount of data to the study. Of the 34,781 activity records used in this study 14,728 (42.3%) are from Cohort 1, 14,894 (42.8%) are from Cohort 2, and 5,159 (14.8%) are from Cohort 3.
Some of the topics at CSU contain several sub learning activities (sub topics) for which data is logged; others only contain an assignment to wrap the topic up at the end. In this exploration, we do not distinguish between the different within-topic models; our analysis deals with engagement with any learning activity inside a topic.

2.3. **Approach**
As this concerns exploratory research, in an iterative process we started with raw student log data from the system and tried to derive metrics that would highlight different kinds of behavior related to the topic tree. We ran analyses using these metrics to find out if we could somehow describe this in terms of effective study behavior. In this paper, we present some of the strongest metrics regarding providing insight and ease of interpretation.

3. **Dimensions of Analysis**
To ease the complexity of the analysis and to aid in the interpretation of student activity patterns, we began by classifying each activity on each of three dimensions.

- **Time** – The time gap between the current activity and the previous activity
- **Distance** – How far the student traveled in the topic tree, measured by the counting edges traversed, and in which direction. Did they stay near the same topic and on the same branch or did they jump to a new branch far away?
- **Purpose** – Why the student attempted this topic. Was it an attempt to Learn the topic or is the student Revising a previously completed topic? We also consider a subscale on this dimension that captures the outcome of any Learn activity. This is a simple flag that indicates if this was the activity on which the student demonstrates mastery and the system marked the topic as complete.

Note that both the Time and Distance dimensions capture the relationship between the current activity and the previous activity whereas the Purpose dimension is concerned only with the current topic and activity. The following sections discuss each of these dimensions individually with later sections detailing the relationship between them.

3.1. **Time**
We begin our look at the time dimension by examining the time of day at which students are beginning activities. A histogram of the activity start times (in the institution’s local time) is given in Figure 3. Here we uncover our first insight - students generally treat their learning as a day job. The bulk of activities (74.5%) take place between 09:00 and 18:00 with the quietest period being from 03:00 to 06:00.
When we consider time in this analysis, we are referring to the time gap (in days) between successive activities. To simplify, we bin this measure into the following discrete categories:

- **SameDay**: the activity starts on the same day as the previous activity
- **NextDay**: the activity starts on the day after the previous activity
- **TwoThreeDays**: the activity starts 2 or 3 days after the previous activity
- **SameWeek**: the activity starts 4 or more days after the previous activity but within 7 days
- **OneWeek**: A gap of 7 or more days but less than 14 days
- **TwoMoreWeeks**: A gap of 14 days or more

In creating these bins, we redefine “midnight” to be at 04:00, the quietest period of the day. This minimizes the potential error of students who study late at night having activities that are part of the same session classified as occurring on different days.

The breakdown of activity by time is given in Figure 4. By far the most common (69.3%) time gap is to complete consecutive activities on the same day. As the time gap increases the frequency of that gap decreases. Students tend to do lots of activities close together. This is an element of the Netflix model of engagement being applied in a learning context. Students are “binge[ing]” on topics leaving short periods between successive activities.

Gaps of two or more weeks are rare. These tend to coincide with breaks that are part of the academic calendar such as mid or end of term breaks.
3.2. Distance
Recall that the topic tree is a directed acyclic graph. While it closely resembles one, the topic tree is not a Tree in the graph theory sense of the word – branches can overlap and have common topics. Student progress through the topics on the tree from left to right, beginning with the foundational prerequisite knowledge before moving on to the more advanced topics to the right.

Our measure of distance captures two aspects, how far the student traveled in the topic tree, measured by the counting edges traversed, and if they moved to a new branch or stayed on the same branch. We visualize students moving between two topics on the tree as involving some movement backward from the first topic to a common prerequisite and then forward to the second topic.

The distance traveled is measured as the number of edges traversed on the tree. Due to two topics potentially sharing many common prerequisites, we take the distance to be the minimum of all possible values. We can define the distance traveled between two topics to be the sum of the distance backward \( B \) and the distance forward \( F \), \( D = B + F \). We can then classify the distance between to topics be Near \( (D < 5) \), Mid \( (5 \leq D < 10) \), or Far \( (D \geq 10) \).

The Topic Tree contains two mandatory orientation topics for all students that are preassigned at the start of the course. These two topics are prerequisites for all other topics in the tree. A student is considered to have moved to a new branch (NewBranch) if the shortest path through the tree, using the backward and forward traverse mentioned above, passes through either of these two orientation topics; otherwise, they have stayed on the same branch. Based on the combination of backward and forward movements, we distinguish between the following subcategories (with examples from Figure 2):

- **Same** - staying on the same topic \( B = 0, F = 0 \) (A->A)
- **Next** – move to a direct post-requisite $B = 0, F = 1$ (A$\rightarrow$C,D)
- **Post** – skipping over the next topics and moving to some other direct post-requisite topic $B = 0, F > 1$ (A$\rightarrow$F,H)
- **Previous** - moving to a direct prerequisite topic $B = 1, F = 0$ (A$\rightarrow$T₁)
- **Prior** - skipping over the previous topics and moving to some other prerequisite $B > 1, F = 0$ (A$\rightarrow$T₀)
- **SameBranch** - moving to some other topic on the same branch $B > 1, F > 1$ (F$\rightarrow$G,J)
- **NewBranch** - moving to some other topic on a different branch, requiring a pathway through $T₁ B > 1, F > 1$ (A$\rightarrow$B,E,G,J)

We can combine the above classifications to create our distance dimension. This is visualized in Figure 5. Note that due to their definitions the Same, Next and Previous classifications can never have total distance traveled of more than one, so labels such as NextMid or SameFar cannot exist. Our distance dimension has 12 categories.

The breakdown of activities on our distance dimension and the percentage of activities based on distance traveled, and movement on branches is provided in Figure 6.

Several interesting patterns emerge when looking at the distance dimension, the most obvious being that in our dataset of 34,781 activities, 15,996 (46%) involve not moving anywhere but instead repeating the same topic. Students repeatedly attempt to learn or revise over several consecutive activities. When looking at the breakdown by total distance traveled we see that it is more common when moving to a new topic to stay near rather than move some distance away. This is the second element of the “Netflix”-style binge on topics we mentioned earlier. Students tend to stay in a small region of the tree and attempt multiple topics with a small-time gap between them.
The next most popular direction is to move to a new branch of the topic tree. When they do this, they are more likely to move a Mid or Far distance away from their previous topic. This happens for many reasons. One possible reason is that a student reaches the end or a far as they wish to progress, of a branch and when moving to a new branch, must traverse a considerable distance back to the root of the tree and forward to the next topic. A second reason is that when a student finishes with a branch, either through choice or from no longer being able to make progress, they want to jump to some other area that is new.

When they stay on the same branch (sBranch), but not do not go to a direct prior (Previous/Prior) or post-requisite (Next/Post), they are approximately equally likely to go any distance. If they do go to a prior or post-requisite, they are far more likely to stay Near the previous topic.

An interesting category to highlight is the nBranchNear. In this case, the students do not travel far but move to a new branch. This can only happen when students are completing activities near the root of the tree, and it could be a sign of student shopping around and deciding on which branch to follow at the start of the course.
3.3. **Purpose**

The final dimension is purpose - the reason why the student is attempting the current learning activity. This dimension has two possible values – Learn or Revise. This dimension captures the cognitive decision being made by the student – what they intend to achieve from the learning activity.

Attached to this dimension is a subscale which captures if the student completed the topic as a result of this learning activity. As stated earlier, a completion is when a student completes the assessment for a topic, either by working through the automated online assessment in topics where available, or by uploading a successful submission for manual marking.

It is worth noting that it is not uncommon for students to be required to resubmit an unsuccessful assessment; the Topic Tree works on a mastery paradigm, and as such a degree of iteration is to be expected. The analysis presented in this paper does not distinguish Revise activities that are in service of a resubmission from those that take place after a successful submission; this may have the effect of confounding some of the analysis that follows.

Table 1 provides a summary of the Purpose dimension and the completion subscale. It displays the frequency, percentage, and the number of unique students for Learn and Revise activities in the data. Interestingly, three of the students never engaged in Revision.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Frequency</th>
<th>Percentage</th>
<th># Students</th>
<th>Completions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learn</td>
<td>31,846</td>
<td>91.6</td>
<td>74</td>
<td>6,764</td>
</tr>
<tr>
<td>Revise</td>
<td>2,935</td>
<td>8.4</td>
<td>71</td>
<td>-</td>
</tr>
</tbody>
</table>

*Table 1: A summary of the Purpose dimension and Completion subscale*

About 1 in 12 activities are revisions. It is not surprising Learn far outweigh Revise in the data as the topics are intended to take approximately 3 hours to complete. We observe that students generally take multiple attempts or sittings to make it all the way through. On the other hand, when revising they tend to do it in one sitting – a student does need to cover all the material in the lesson and can concentrate on the areas they most need to review.

The table also provides the total number of the Learn activities that were also flagged as completions. The completion flag accounts for 21.2% of all Learn activities, suggesting that it takes students on average around five interactions before they are ready to submit for assessment for a task.

4. **Relationship Between Dimensions**

4.1. **Distance Versus Time**

Figure 7 shows the breakdown of Distance versus Time dimensions - note that the fill color is on a log scale. The most obvious stand out is the high concentration of students on attempting the same topic on the same day. This is in line with earlier findings – students are bingeing on topics in a local area of the tree with a short period between topics (they happen on the same day). We also see two other features jump out, to move to a far pre- or post-requisite (FarPrior/FarPost) is rare regardless of the time gap. The patterns of distance vs time interactions are consistent across the three student cohorts; as such the data in Figure 7 are presented in overall aggregated form.
To get a better understanding of what is happening in Figure 7, we provide simplified alternative views in Figure 8. Here we simplify the distance dimension to its two aspects. Part (a) visualizes the Time dimensions versus movement on the branches. Part (b) shows the Time dimension versus distance traveled. In both cases, the y-axis shows the percentage of activities for each category for a given time gap.

Figure 8(a), shows that when a student completes two consecutive activities on the same day, there is an over 60% chance that this student will attempt the same topic on both activities. This is three times more likely than for a student to move to a new branch, which is the case for just over 20% of all consecutive activities on the same day.

If a student waits until the following day to attempt the next activity, then the likelihood of attempting the same topic is approximately equally likely as attempting a topic on a new branch. As the time gap between activities grows students become more likely to move to a new branch than anywhere else.

Figure 8(b) shows the same pattern, as time increases repeating the same topic becomes less likely. Now we see that moving some distance away becomes more likely as the time gap increases but the exact distance traveled is not dependent on time. This can be inferred as with Near, Mid and Far are approximately equal. This breaks down as time gap grows: if a student leaves a gap of two or more weeks, they are more likely to go further away from their last topic.
4.2. Purpose Versus Time and Distance

There are two possible values for Purpose – Learn or Revise. To highlight the most interesting findings in the most straightforward manner, we focus on just the Revise activities. For each of the categories on the Time (Figure 9) and Distance (Figure 10)
dimensions, we will analyze and display the proportion of all activities in that category that are revisions.

If a student does a second activity on the same days as the previous activity, there is an almost 10% chance that that activity will be a revision – Figure 9. It appears that while bingeing on topics on the same day, students are regularly sprinkling in revision of previously mastered topics. When students do not complete another activity until the next day, the percentage of activities that are revisions halves. As the time gap grows the rate of revision does increase, although there is a small drop if the gap is two or more weeks. Generally, the more significant the gap, the more likely it is that the student will start with a revision, although they are still more likely, for every time gap, to begin with a Learn.

When examining this the percentage of the actions at each distance that are revise in Figure 10, we see some interesting patterns emerge. First, if a student moves to a prior topic, that is a topic that is a direct prerequisite of the previous topic, they are far more likely to engage in a revision than if they went to any other topic. In fact, the further back the prior topic the more likely the student is going there to revise.

![Figure 9: The percentage of activities that are Revise for each category on the Time dimension.](image)

The distances with the lowest rate of revision are nBranchNear or Next. This makes sense given the underlying meaning of these distances. If the move to the next topic they are more than likely progressing through the tree learning and mastering topics, with all post-requisites of the current topics still to be learned. As stated earlier, moving to a nBranchNear can only occur near the root of the tree. Students will generally be in this area of the tree at the start of the programme and will therefore not be engaged in as much revision.
5. Conclusion

In this paper we were interested in understanding better how the students behave in this online environment. The CSU’s curriculum is unique and learning to understand the students’ behaviour is a first step to understanding what behaviours lead to academic achievement in the programme. In previous work we studied how students move through the topic tree using total distance traveled on the tree and time gap between activities. This work was enlightening as we established that students do a considerable amount of revision, but we did not distinguish between the various directions students took. In this work, we extended the analysis to consider the distance traveled, the direction through the tree and the purpose of the activity.

What we found is that most students attempt the same topic multiple times on the same day. If the next activity takes place on the next day, revision is as likely as starting a new topic. We found that students need on average around five activities before they complete a topic successfully and we found that the proportion of Learn activities versus Revise activities is 91 to 9 per cent. Most revisions concern topics that are direct prerequisites to the activity on the same branch of the topic tree. This indicates that students tend to finish and revise within a branch before they move to the next branch on the tree.

Figure 10: The percentage of activities that are Revise for each category on the Distance dimension along with the same for each aspect of this dimension.
These insights shed a new light on the binge behaviour we observed in earlier work. Students tend to stay on the same activity and the same branch and often repeat the same ‘episode’ before moving on, but they do not just move forward. They also review previous topics before they attempt something new within the same ‘series’ and they often go back in the ‘series’ to check out parts of ‘episodes’ that matter for their understanding of the full series.

This more detailed analysis of the patterns of topic acquisition has somewhat undermined the previous “Netflix”-style understanding of student progress as being too simple a model to explain the various ways in which our cohorts learn, and not taking into account the differences in students’ patterns of engagement with entertainment vs education. Future work will allow us to expand the model to capture these nuances more fully.

6. References


