

---

## THE STUDENT STUDY EXPERIENCE – ANALYSING STUDENT STUDY CHOICES

*Chris Edwards, Mark Gaved, The Open University, United Kingdom*

---

### Abstract

As higher education institutions increasingly teach online and offer greater levels of choice to students (over which modules to study, in which order to study, and how long to extend study before qualification) new challenges are introduced. One of these challenges is how to maintain an understanding of the student experience. This understanding is necessary to provide feedback to both students and faculty, and institutionally for the continued enhancement of quality. This paper is the first attempt at providing a narrative describing one approach to this challenge and the experience within a large distance learning University. It demonstrates a new approach to data is key to enabling the analysis of student study pathways. For many years, this University has offered great flexibility of study and as wide a study choice as it is possible to offer with conventional modules. By design, the Institution holds high levels of data for all student study. However, whilst it is possible to create bespoke queries, we found that this has been insufficient to readily enable analysis of the student experience. By moving from a traditional relational database structure to a multi-model database, many of the difficulties are resolved. In this paper, we report on this approach and describe next steps, including the potential to apply machine learning algorithms and test other data theories like that of Markov Chains.

### The need for a different approach

Colleagues within the Institution have for years asked questions relating to student study paths because they recognise this is central to understanding whether there are any issues with particular paths, or recommendations that could be made to students to aid them in their choices. Whilst some of the simpler questions have been taken up the level of resource required has meant that mostly these questions have remained unanswered. Within the last decade there has been one major change in the way higher education operates within England. The loss of government funding resulted in the introduction of student loans and forced universities to increase fees. This change also had the effect of increasing students' expectations of universities. To qualify for a loan, students are required to register on a

degree course: an entire programme of study rather than an individual module. Together these changes greatly increased the need to improve our understanding of student study choices. However, after several years of attempts, there had been little progress. The problem was proving to be intractable. The authors postulated that the issue was not to do with the expertise within staff but much more likely to be to do with the way data was organised. The University had developed its datasets gradually over many years, and these are held in traditional relational databases. It should be said that these fully support the day to day running of the Institution but are inefficient at supporting queries relating to student study paths. Recent advances in database technology have produced so-called 'graph' databases that are tuned to manage linked data.

### ***Fragmentation of cohort***

When considering student performance on almost any of the University's modules, it is clear that the cohort is fragmented. Yet, we have struggled to find a way of dealing with this fragmentation in our analyses. This is actually the same challenge of analysing study pathway but from the module rather than student perspective. This fragmentation is created by several factors:

- The programme of study leading to a qualification offers choice at least at one point; where students can opt to study either module A or module B and still complete the qualification. One of the most constrained programmes of study, because of the requirements of professional bodies is Psychology. However, by the time a student reaches their final module they could have taken one of sixty-four different study paths. The Modern Languages degree has rather more options for study and a student completing just the first of three study stages will select from some 1436 study paths.
- Students can choose how long they take to complete a degree. This usually ranges from three years, the equivalent of full-time study, to a maximum of 16 years. Therefore, a cohort of new students embarking on the first module of a programme of study will soon find themselves split up, even if there is no optional study choice within the programme. This is simply because some will study three modules per year and others will choose to take either one a year, or to take a study gap between modules.
- There is also fragmentation due to study outcome. Ideally, we would like all students to be successful at their first attempt but we know this does not always work out. The most obvious reason is that some students complete their study but do not pass the module. Others may realise part way through that they need to halt their study and return to it at a later date. These defer their study. Still, others

withdraw from one module, perhaps because they realise it was the wrong choice for them, and move on to a different one.

- Many modules serve a number of qualifications. So, at any one time, Module A will be studied by students on different study paths to Qualification A and Qualification B, and so on.

Any consideration of the effectiveness of a module or programme of study therefore, needs to be able to recognise these fragmentations and be able to separate them out as necessary for a fuller understanding to be gained. If we do not do this, we are not using the available data to the full and missing potentially important lessons.

### **The student pathway data model**

We selected a multi-model database, ArangoDB. This is a flexible technology that fully supports graph database structures and queries, whilst also allowing other approaches to be used without impacting on the data. Having made this selection, the next steps were around determining how the data should be structured to facilitate pathway queries. The only requirement we set was for the model to support pathway analysis. However, the choice of database should also futureproof the model, allowing it to be readily adapted to any future needs. At the outset it was important to define what we have called the *smallest creditable element*. For our institution, at this moment in time, this element is the combination of student, module, and module presentation date: student-module-presentation. A student could choose to study Module A in 2020. There are usually several potential outcomes for this study attempt: pass, fail, defer, and withdraw. If the student passes, they may go on to study Module B in 2021, creating a new data element. If they fail, they may re-attempt, and this new study attempt would create another data element.

### **Hypothesis**

We hypothesised (Edwards, 2017) that within this database, two complementary pairs of data structures would essentially provide all that was needed to construct student pathways and analyse them in terms of qualification programmes of study. The four data structures are shown in Figure 1 below. They comprise a set of vertices containing all student study attempts; a set of edges linking one study attempt with the next; another set of vertices containing all modules; a set of edges linking the modules into qualification programmes of study.

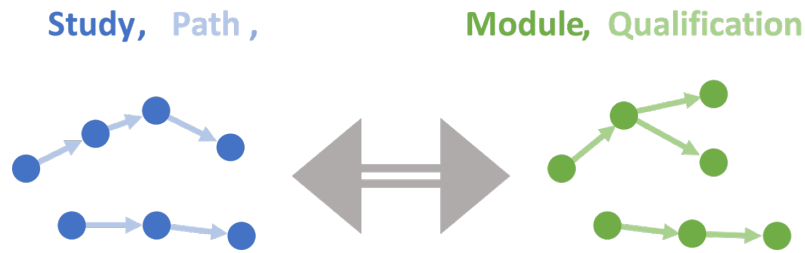


Figure 1. The four data structures within the model

## The model in practice

Early results from using this model are very encouraging. For example, an analysis of the actual study paths followed by a single starting cohort of students is shown in Figure 2 below. Some of the data in this figure has been deleted in order to anonymise the figure. The figure was produced by selecting all the new students beginning their study with a defined module and building all their individual study paths from that point. The query then aggregates the data and outputs it in a form the database interprets as a graph. This graph could be designed to be interactive, not just with the ability to move the vertices around to improve visibility but also in enabling the expansion of individual vertices to explore the data in greater detail. The query can be readily modified to focus on a different module, or group of students. This kind of analysis was presented in both Ullmann et al. (2018) and Clow et al. (2019).

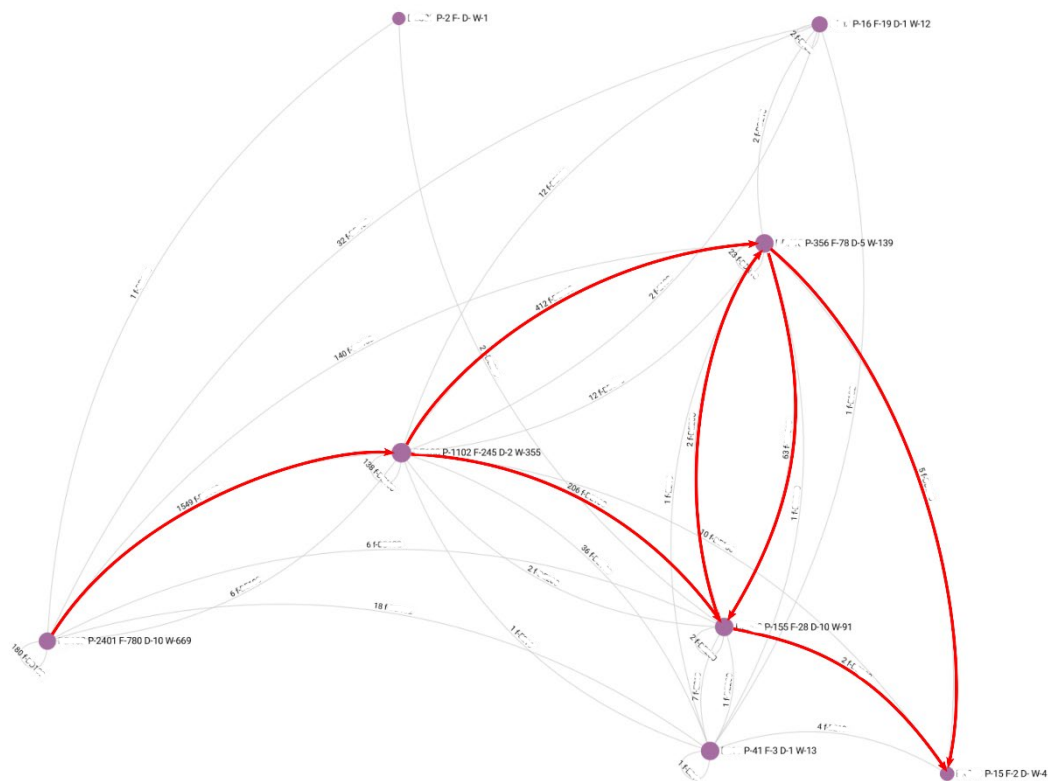


Figure 2. The study paths followed by an initial cohort of new students studying an undergraduate degree programme. The figure shows numbers of pass, fail, defer and withdraw for each study step. The most populace paths are highlighted in red. Note, the diagram has been anonymised through the deletion of some data.

Another example of the model in use is the simple representation of study paths. For this a simple standardised form of representation was created. It is as follows, an arrow,  $\rightarrow$ , is used to signify that study of one module was started after that of another module. Brackets are used to indicate that two or more modules are studied concurrently. These modules are always listed in alphabetical order. Therefore, the study path of a student studying Module A, followed by Module B and Module C studied concurrently, and finally, Module D, would be written as:

Module A  $\rightarrow$  (Module B, Module C)  $\rightarrow$  Module D

Data queries were developed that build the study path for a defined group of students and then aggregate the results for further analysis. This could be readily carried out within any spreadsheet programme. The queries are readily updated to select any student group and the standardised output. Table 1 below contains anonymised data from one real query in the faculty of Science, Technology, Engineering and Mathematics, STEM. By developing this standardised representation, colleagues without knowledge of the database or query language can be provided with data that they can readily work with, using tools with which they are very familiar.

**The Student Study Experience – Analysing Student Study Choices**

Table 1: An anonymised sample of output of a study path analysis, presented in standardised form. M, indicates module; Y, year; P, Pass; Ppath, Pass path; Ps, Passes. The entries have been abbreviated and the column showing module attempts removed, in order to fit on the page.

Fullpath	Ppath	Ps(a-to-z)	credit	freq
(M1-Y1-P,M2-Y1-P)->(M3-Y2-P)->(M4-Y3-P)	(M1,M2)->(M3)->(M4)	M1,M2,M4,M3	120	57
(M1-Y1-P,M2-Y1-P,M3-Y1-P)->(M4-Y3-P)	(M1,M2,M3)->(M4)	M1,M2,M4,M3	120	19
(M2-Y1-P,M3-Y1-P)->(M1-Y2-P)->(M4-Y3-P)	(M2,M3)->(M1)->(M4)	M1,M2,M4,M3	120	18

The data model readily supports investigations into the health of a qualification through two different methods. One of these methods is through analysing the success of those students who have attempted to study the required modules, in the correct order: if there is one. This can be done, both in terms of overall outcome and in terms of grade for those passing. The other method that we have tested, is that of taking a snapshot of the qualification. This method considers all students anywhere on the programme of study for the qualification at a certain point in time, considers their study journeys to that point and their ongoing study. Each of these methods has its own strengths and when used together they can complement each other to build a rich picture of the qualification and the student experience. To date, both these methods have been trialled on one qualification in the Social Sciences.

## Next steps

As described, this approach to the data has shown much promise and is beginning to allow longstanding questions to finally be answered within the institution. We are embarking on a University project to use this technology to drive a qualification dashboard providing standardised information for colleagues. The database also has some built-in machine learning queries which we intend on applying to the data. This may yield some patterns that are not readily seen through the usual analyses that we undertake. We are finalising a sub-project that took a first step towards exploring whether the probability theory of Markov Chains, has any relevance to University study, after finding a paper that had attempted something similar (Ikonen, 2009). The theory is used in a number of fields, including voice recognition. Whilst this fact may not immediately suggest any value for universities, it seems that from our initial efforts it may well be relevant to, at least, some groups of students. This is currently being checked.

## Conclusion

By responding to the long standing challenge of analysing student study paths within one institution through structuring data in a different way using database technologies that are still relatively new, we are finding that many questions that have long gone unanswered are finally yielding. We have also found that the model developed has potential to serve the Institution more widely through the development of a qualification dashboard. In addition,

this restructuring of data will enable machine learning algorithms to work with the potential to reveal new insights. Other theories, like that of Markov Chains can also be tested against the data. These developments amount to a significant step forward for our Institution and its potential to support students. As noted in the abstract, many institutions are moving more of their teaching online and this data model may also therefore offer benefits more widely.

## References

- Clow, D., Coughlan, T., Cross, S.; Edwards, C., Gaved, M., Herodotou, C., Nguyen, Q., Rienties, B., Thorne, S., & Ullmann, T. (2019). *Scholarly insight Winter 2019: a Data wrangler perspective*. Milton Keynes: Open University. Retrieved from <http://oro.open.ac.uk/59646/>
- Edwards, C. (2017). *Understanding student experience in the age of personalised study*. Paper presented at the ALT Online Winter Conference 2017, 12-13 Dec 2017. Retrieved from <http://oro.open.ac.uk/53745/>
- Edwards, C., & Gaved, M. (2019). *Analysing student study choices to qualification: which study routes are best?* Paper presented at the ALT Annual Conference 2019, 3-5 Sep 2019, Edinburgh. Retrieved from <http://oro.open.ac.uk/66607/>
- Ikonen, E., Tervaskanto, M., & Najim, K. (2009). Analysis of students' study paths using finite Markov chains. *Proceedings of the 2009 IEEE International Conference on Control and Automation, ICCA 2009*, 1813–1818. <https://doi.org/10.1109/ICCA.2009.5410623>
- Ullmann, T., Lay, S., Cross, S., Edwards, C., Gaved, M., Jones, E., Hidalgo, R., Evans, G., Lowe, S., Calder, K., Clow, D., Coughlan, T., Herodotou, C., Mangafa, C., & Rienties, B. (2018). *Scholarly insight Spring 2018: a Data wrangler perspective*. Milton Keynes: Open University UK. Retrieved from <http://oro.open.ac.uk/56732/>