USING ABSORBING MARKOV CHAINS TO ASSESS THE PROGRESS OF DOCTORAL AND MASTER DEGREE CANDIDATES

Miles G Nicholls, Faculty of Business and Enterprise, Swinburne University of Technology, PO Box 218, Hawthorn, 3122, Victoria, Australia, mnicholls@swin.edu.au

ABSTRACT

In this paper, absorbing markov chains are used to analyse the flows of higher degree by research candidates (doctoral and master) within a faculty of business. The candidates are analysed according to whether they are full or part time. The need for such analysis stemmed from what appeared to be a rather poor completion rate (as reported by the University). However, this reported completion rate was a 'macro' figure that aggregated PhD and Master (by full time and part time) completions together. If there really was a problem, then the Faculty needed to know where in the system it was and what potentially might then be done to remedy the problem.

INTRODUCTION

The problem of understanding and assessing the flows of students through educational systems has long been dealt with using finite absorbing markov chains. Early work in this area concentrated on the flow of student teachers at the undergraduate level [1]and flows of students at undergraduate and graduate level within a business faculty [2] respectively. Subsequently, there appears to have been little work undertaken in this specific area. However, most recently, Shah and Burke (1999) published an absorbing markov chain analysis of undergraduate students in the Australian higher education system. However, apart from [5], there still appears to be little work done at the *graduate* level. The driving force behind the study undertaken in this paper, was the need to understand the underlying flow pattern and emerging attributes of a specific faculty's PhD and Master students

THE MARKOV STUDENT FLOW MODEL

The exact same modelling approach was used for both doctoral and master students, thus a generic finite absorbing markov chain model is illustrated below. While the model can be found in a multitude of texts (see for example the classic texts of [1] and [4]), terminology, forms of expression and symbolism vary considerably, and it is necessary define some terms used in this study to remove any confusion.

The expressions (statistics) that permit the estimation of the system attributes are summarized below.

Note:

$$\{\mu_{i,j}\} = (I-Q)^{-1} \qquad (1)$$

(i) Probability of Absorption (Steady State) ($Pr(A_{i,j^*})$) for an entity currently in state i eventually absorbing into state j*. Note here, $P = \{ p_{i,j^*}\}$ where P is a (I X [J*-J]) matrix and represents the probability of moving from a transient state to an absorbing state. $Pr(A_{i,j^*})$ is calculated thus:

$$Pr(A_{i,j^*}) = \sum_{j=1}^{J} \mu_{i,j}\, p_{j,j^*} \qquad (i=1, . . ,I) \quad (j^*=J+1, . . ,J^*) \qquad (2)$$

(ii) First Passage Times (FPT) can be calculated. FPT are expected times that an entrant in state i will stay within the system prior to absorption into any of the (J*-J) absorbing states (an entrant can be an entity that commences in state i or passes into it (if feasible) during its time spent in the system). Q is the matrix of probabilities of moving from transient state i to transient state j, effectively:

$$Q = \{q_{i,j}\} \qquad (i=1, . . ,I) \qquad (3)$$

(j=1, . . ,J). FPTs are calculated thus:

$$FPT_i = \sum_{j=1}^{J} \mu_{i,j} \qquad (i=1, . . ,I) \qquad (4)$$

(iii) Expected Completion Rates (ECR) for an entity having started in state i and absorbing into state j* *in exactly time* t. This is calculated via:

$$ECR^{t^*}_{i,j^*} = \sum_{j=1}^{J} qi,j^{(t^*-1)}\, p_{j,j^*} \qquad \begin{matrix}(i=1, . . ,I)\\(j^*=J+1, . . ,J^*)\\(t=t^*, . . ,T)\end{matrix} \qquad (5)$$

Note here that i will be set to specific values to correspond to common entry states for students in each of the models (PhD and Master) and the absorption states will also be focused on ‘thesis acceptance’ only. The value for t* will be determined by the earliest possible time the PhD or Master regulations permit completion.

(iv) Expected Completion Rates (CECR) for an entity having started in state i and absorbed into state j* *within* time t. This is calculated via:

$$CECR^{t^*}_{i,j^*} = \sum_{t=t^*}^{T} ECR^{t^*}_{i,j^*} \qquad \begin{matrix}(i=1, . . ,I)\\(j^*=J+1, . . ,J^*)\\(t=t^*, . . ,T)\end{matrix} \qquad (6)$$

THE DEVELOPMENT OF THE STUDENT FLOW MODELS

In this section, the PhD student flow model is developed (the Master student model is available from the author. The data set used in this study is a limited one since the relevant school came into existence only some seven years ago, and data relating to candidates preceding this date (who are included where possible since they subsequently ‘transferred into’ the new Faculty) is often corrupted and difficult to reassemble. Consequently, only some data was used from this subset. With the somewhat limited data, some transient states have not yet been used by the student flows. However, sufficient data was able to ensure results were not meaningless.

Raw data in this study consisted of progress records with the *transitions* recorded onto a database using a transition period of one year. The time frame of one year was necessary since the number of entities that could be fully tracked through the system (25 over a period of ten years) was small. Reducing the transition period further would lead to inaccuracy as well as data instability due to a very sparse fundamental matrix (Q). This is discussed later in the paper. While the data set is small, it was relatively accurate. Leave of absence was ignored and transitions allowed to continue sequentially. The University and the Government does not count leave of absence in any duration calculations. The estimates of $q_{i,j}$ and $p_{i,j*}$ values were found using maximum likelihood estimates

It should be remembered that these student flow models have been developed to give some *general appreciation* of the performance of the systems, not provide precise assessment. With the limited size of the database, tests to ascertain the degree of stability of the data are not undertaken. As [2] suggests, high levels of accuracy are not needed nor sought. It is simply not possible as pointed out above to obtain highly accurate data and/or to incorporate all of its components into the flow models.

The Doctoral Student Flow Model

Full time doctoral candidates within the University generally cannot complete their degree before the expiration of two years but must complete it within five years. For part time candidates, the minimum time is three years and the maximum time eight years. Transference from part time to full time and *vice versa* is permitted. Examples of the transient states definitions are: i=1 :Beginning of Year 1 - full time, i=2 :Beginning of Year 2 – full time and i=6 :Beginning of Year 1 - part time. For absorbing states, definitions are along the lines: j*=14 and j*=15 representing Withdrawal and Thesis Accepted respectively. At the expiry of the maximum duration times as specified by the relevant PhD and Master Regulations, students are automatically removed from 'enrolled' status to 'not enrolled' status and classified as a 'Withdrawal'. Approximately 25 PhD candidates were tracked over 64 transitions from 1994 to 2004.

RESULTS

The steady state probabilities of absorption as calculated as per (2) show that for a PhD candidate starting in the first year on a full time basis (i=1), that the probability of eventually having a thesis accepted is 0.6528 (i.e., 65.28% of candidates will have a thesis accepted having commenced full time year 1), while for a candidate commencing year 1 on a part time basis (i=6), the probability of having a thesis accepted is 0.4075. The implication from the above is that the probability of *withdrawal* given a candidate *commences* full time and part time is 0.3472 and 0.5925 respectively.

The First Passage Times (FPT as calculated in (4)) prior to absorption indicate that on average, a PhD candidate commencing in full time Year 1 category, will be in the 'system' on average 3.9745 years before absorption (i.e., into 'thesis accepted' or 'withdrawal'). Most withdrawals occurred at maximum candidature time, therefore this figure is expected to reflect completion times quite accurately. The FPT allow for changes in status, i.e., from full time to part time and if appropriate back again to full time *ad nauseum*. Part time candidates commencing in year 1 can on average expect to spend 3.0481 years in the system prior to absorption (in 'thesis accepted' or 'withdrawal'). For full time candidates, this average time coincides with the duration for which they are funded by either the University or the Commonwealth Government. The part time candidates' average duration in the system is more likely to involve withdrawal than for full time candidates and this will be likely to happen after a shorter duration in the system that for full time candidates.

The expected completion rates for doctoral students within *exactly* a specific number of years (as in (5)) in the system indicates that the duration after having entered the system where the greatest probability of successful completion occurs is after 4 years for full time candidates (0.4444) and after 5 years for part time candidates (0.2963). The former coincides the with the first passage times, confirming also the pattern associated with the steady state probabilities of completion. However, the more useful and meaningful of ‘completion’ rates will be the cumulative sum of the exact rates (probabilities) as detailed in (6), i.e., the *within* completion rates.

The exact rates indicate that expected completion rates after five years for Ph D candidates commencing in full time and part time year 1 categories are 65.27% and 40.74% respectively and that the maximum number of expected completions for full time candidates occurs within 4 years while for part time it is after five years (allowing for the fact that candidates could have become part time having commenced as full time ones).

CONCLUSIONS

The part time completion rate for PhD Candidates is a concern (as it is approximately 41%). While this figure to some extent has an explanation (i.e., the completion of a part time degree is fraught with many more risks mitigating against completion than a full time candidature and the number of completions to date small because of the time scale [i.e., part time candidates generally not through the system yet]), still the completion rate seems disappointingly low. The completion rate for full time PhD candidates benchmarks very favourably with other faculties and universities in Australia

REFERENCES

[1] Bartholomew, D. J. *Stochastic Models for Social Processes*’, Second Edition, John Wiley and Sons, London, New York, Sydney and Toronto, 1973.

[2] Bartholomew, D. J. and Forbes, A. F. *Statistical Techniques for Manpower Planning*, Chichester, J. Wiley and Sons, 1979, p107.

[3] Burke, G., *A Study in the Economics of Education with Particular Reference to the Supply of Secondary Teachers for Government Schools in Victoria*, PhD Thesis, Monash University, 1972.

[4] Kemeny, J. G. and Snell, J. L. *Finite Markov Chains*, The University Series in Undergraduate Mathematics, Eds., Kelley, J. L and Halmos, P. R.., D. Van Nostrand and Company, Inc., Princeton, New Jersey, Toronto, London, 1967.

[5] Nicholls, M. G. A Markov Evaluation of a Tertiary Education Faculty, *Higher Education*, 1983,12, pp 721 – 730.

[6] Shah, C. and Burke, G. An Undergraduate Student Flow Model: Australian Higher Education, *Higher Education*, 1999, 37, (4), pp 359 – 375.