Network Analysis of University Courses

Ahmad Slim
University of New Mexico
Albuquerque, NM 87131
ahslim@unm.edu

Jarred Kozlick
University of New Mexico
Albuquerque, NM 87131
jkozlick@ece.unm.edu

Gregory L. Heileman
University of New Mexico
Albuquerque, NM 87131
heileman@unm.edu

Jeff Wigdahl
University of New Mexico
Albuquerque, NM 87131
jwigdahl@ece.unm.edu

Chaouki T. Abdallah
University of New Mexico
Albuquerque, NM 87131
chaouki@unm.edu

ABSTRACT

Crucial courses have a high impact on students progress at universities and ultimately on graduation rates. Detecting such courses should therefore be a major focus of decision makers at universities. Based on complex network analysis and graph theory, this paper proposes a new framework to not only detect such courses, but also quantify their cruciality. The experimental results conducted using data from the University of New Mexico (UNM) show that the distribution of course cruciality follows a power law distribution. The results also show that the ten most crucial courses at UNM are all in mathematics. Applications of the proposed framework are extended to study the complexity of curricula within colleges, which leads to a consideration of the creation of optimal curricula. Optimal curricula along with the earned letter grades of the courses are further exploited to analyze the student progress. This work is important as it presents a robust framework to ensure the ease of flow of students through curricula with the goal of improving a university's graduation rate.

Categories and Subject Descriptors

G.2.2 [**Graph Theory**]: Network flows

General Terms

Algorithms, Experimentation

Keywords

Complex networks; longest path; curriculium complexity; cruciality; student progress

1. INTRODUCTION

Many definitions of student success exist in the literature. While these vary from grades and persistence to self-improvement, most studies consider graduation the ultimate measure of student success [7]. For a college student, having a bachelor's degree has become a necessity with attainment rates topping 30% for adults over the age of 25 according the latest census numbers [1]. From the university's perspective, and especially for public universities, the definition of student success broadens from graduation into student retention rates and time-to-degree. These factors are important because many states have tied a percentage of the university's funding directly to such student success metrics [2]. This so-called "performance funding" has become a popular way to incentive universities to help students graduate in a timely fashion. Whether a causal relationship exists between performance funding and graduation rates remains to be seen, but studies have clearly shown a rise in graduation rates as state appropriations per student increase [8].

There are many factors that correlate to a student graduating. These factors may be broken down into two categories, pre-institutional and institutional factors. Some pre-institutional factors include high school GPA and socioeconomic data, whereas institutional factors include the specific institutional processes and policies that help students graduate while minimizing or even excluding the student's actions [5]. Institutional factors contributing to success include any quantifiable actions that occur during the student's time at an institution. Studies have identified the degree to which factors such as learning centers, freshman year programs, dorms, study rooms, etc. contribute to student success [6].

There is one institutional factor that is often overlooked; however, which is the curriculum associated with a particular degree program. Accordingly, in this paper we investigate the structure of university course networks at three different levels: department, college and university. Due to the nature of course networks, we use graph theory and complex network analysis to provide a mathematical foundation for detecting crucial courses and hence analyzing the important features of course networks which may help administrators to make decisions on when to offer certain classes, who should teach them, and what is truly necessary for a degree in a certain field.

The remainder of this paper is organized as follows. We introduce a network model in Section 2, and factors and parameters used to study and quantify the cruciality of courses in the network are described in Section 3. Section 4 presents a case study that includes numerical results for the University of New Mexico (UNM) course network. Applications

WWW'14 Companion, April 7-11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579360.

utilizing the course cruciality framework are described in Section 5. Finally, Section 6 presents some concluding remarks.

2. CONSTRUCTION OF THE COURSES' NETWORK MODEL

Using graph theory as the basic method to study curriculum graphs, we first build a model for the curriculum graph structure, and then generate the graph model by abstracting the courses into nodes and connecting two nodes with a directed edge if there is a pre-requisite relationship between the courses associated with the nodes. The course network is therefore represented as an $N \times N$ adjacency matrix M, where N is the number of nodes in the network. If there is an edge from node j to node i, then $M_{ij} = 1$, otherwise $M_{ij} = 0$.

3. ANALYSIS FRAMEWORK FOR THE COURSES' NETWORK

The cruciality of a course within a network is related to two main features, its delay factor and its blocking factor, and these two factors are characterized by two additional parameters, the longest path and the connectivity. The longest path L_i of node i is defined as the length of the longest path passing through node i. The connectivity, V_i, of a node i is defined as the total number of nodes connected to i. That is, n_{ij} is 1 if there is a path from i to j and 0 if no such a path exists. Then the connectivity V_i is given by

$$V_i = \sum_j n_{ij} \tag{1}$$

Examples illustrating the significance of these parameters in quantifying the cruciality of courses are provided below.

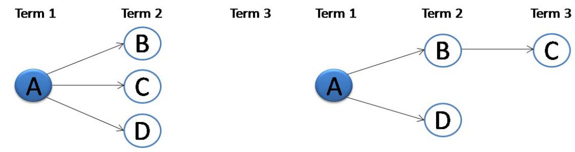
3.1 Delay Factor

Some courses have a critical impact on the academic progress of a student in the sense that any failure in these courses (or delays in taking them at the appropriate time) subjects the student to the risk of not finishing on time. It is therefore essential to detect these courses. The following example illustrates a process for detecting them using longest path length parameter.

Given four nodes A, B, C and D representing four different courses, possible relationships between are shown in two different scenarios in Fig. 1. In Fig. 1(a) course A is the pre-requisite of B, C and D, while Fig. 1(b) shows the same courses, but with different prerequisite relationships between them. In the latter, A is the prerequisite of B and D whereas B is the prerequisite of C. Comparing these two figures, it is clear that A in Fig. 1(b) is more crucial than it is in Fig. 1(a). This may be explained as follows. Assuming a three-term curriculum, a student who fails A in Fig. 1(a) still has the chance of finishing on time, whereas one who fails A in Fig. 1(b) ends up with more than three terms and thus is delayed. This phenomenon is reflected by the length of the longest path,L_i. In Fig. 1(a), the longest path value of A is one whereas in Fig. 1(b) it is two. However, the value of the connectivity for A is three in both scenarios.

3.2 Blocking Factor

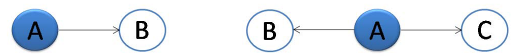
In addition to the delay factor, it is natural to conclude that a course that is a prerequisite to a large number of other



(a) Node A is pre-requisite to nodes B, C and D

(b) Node A is prerequisite to nodes B and C whereas B is prerequisite to C

Figure 1: Graphs illustrating course cruciality using the longest path length factor.



(a) Node A is pre-requisite to node B

(b) Node A is pre-requisite to nodes B and C

Figure 2: The two graphs illustrate the cruciality of node A using connectivity factor.

courses is more crucial. If a student fails such a course or does not attempt and pass it at the right time, the student may be blocked from attempting follow-on courses, leading to a negative impact on progress. This is illustrated by the following example. Nodes in Fig. 2 represent three different courses. Nodes in Fig. 2(a) are linked differently than those in Fig. 2(b). Node A in Fig. 2(a) is a prerequisite to node B whereas in Fig. 2(b) it is a prerequisite to nodes B and C. Comparing these two figures, it would be reasonable to consider node A in Fig. 2(b) more crucial than it is in Fig. 2(a). In the case of failure or delay, node A in Fig. 2(b) will block more courses. This result is reflected by the value of the connectivity, V_i. In Fig. 2(a), the connectivity of A is one whereas in Fig. 2(b) it is two. However, the value of the longest path length for A is one in both scenarios.

Note that other parameters such as in-degree and out-degree measures are not as suitable as the longest path and connectivity parameters. For example, if we consider the in-degree and out-degree parameters instead of the longest path length parameter to compute the crucaility of node C in Fig. 1(a) and Fig. 1(b), both scenarios would lead to the same crucaility value which does not differentiate between the two scenarios despite the fact that node C in Fig. 1(b) is more crucial than that in Fig. 1(a) taking into consideration the delay factor discussed previously.

Based on the foregoing discussion, the cruciality of course i, denoted C_i, is defined as follows:

$$C_i = V_i + L_i \tag{2}$$

Note that Eq. (2) does not take into consideration any prior knowledge of the difficulty of a course which might be reflected by the failure rate of the course. Though such knowledge may be a critical factor to include in measuring cruciality Eq. (2), it is not taken into consideration in this work.

4. NUMERICAL RESULTS

In order to empirically validate our proposed cruciality metric, we analyzed actual university data provided by UNM.[1]

4.1 Cruciality Analysis of UNM Courses

In determining the most crucial courses at UNM, all UNM courses with their respective pre-requisite links were mathematically represented by an adjacency matrix M.

4.1.1 Data Pre-processing

It is well known that building relationships between courses based on pre-requisite links is not trivial. For example a course i may be a co- or pre-requisite to another course j and/or vise-versa. In order to deal with such relationships, some assumptions were made:

1. If course i is a co- or pre-requisite to course j, we assume that i is a pre-requisite to j. In other words, we assume the worst case scenario where course i and j cannot be taken in the same academic term.

2. If course i is a co- or pre-requisite to j and vice-versa or in other words if courses i and j are co-requisites, we consider the worst case scenario in which one of the courses is considered to be the pre-requisite of the other. In this case we eliminate cycles from our graph.

Besides the assumptions made for co- and pre-requisite relationships, it is also important to consider the transitivity relationship links between courses. These types of links must be deleted. For example if course A is a pre-requisite to both courses B and C, while C itself is a pre-requisite to B, then there is no need to show that A is pre-requisite to B. Otherwise, A assumes extra cruciality that is not deserved. We use a Transitive Reduction algorithm to filter out the transitive edges within M [3]. Note that the cruciality value is set to 1 as initial value for all the nodes.

4.1.2 Basic Analysis

Next, we show the cruciality distribution for all 7593 courses at UNM, along with the 10 most crucial courses. We evaluated the cruciality for each course based on Eq. (2). Figure 3 shows the cumulative distribution function for the cruciality of UNM courses created using the method described in [4]. Based on the evidence of Fig. 3, it is shown that the cruciality distribution follows a power law with an alpha value of 2.1. In other words, Fig. 3 shows that the number of courses with high cruciality is much smaller than those with low cruciality. In addition, Table 1 shows that the 10 most crucial courses at UNM are mathematics courses. These courses are labeled as highly crucial courses by UNM. Based on some basic statistical results, UNM's analysts have already concluded that the graduation rate is negatively influenced by these particular courses. This analysis and conclusions match our predicted results using the cruciality method.

4.2 Comparison of Curricula Complexity within a College and Across Colleges

In this section we study the complexity of a particular curriculum and compare it to another curriculum within a college or across other colleges. For example, the electrical engineering curriculum's complexity is compared to that of the computer engineering curriculum. It may be also beneficial to examine if the engineering school's curricula, on average, are more or less complex than the business school's. We thus examined the complexity of curricula within two colleges at UNM, the School of Engineering (SOE) and the Anderson School of Management (ASM). For each program within these colleges, we created a network to represent its curriculum. This network has nodes to represent courses within the program's curriculum, and edges to model prerequisite relationships among courses. We thus created 9 different networks for the SOE representing 9 different programs and 11 different networks for those representing ASM. For each program's curriculum we calculated its complexity defined as the cruciality sum of all the courses within the curriculum:

$$S = \sum_{i}^{n} C_i \tag{3}$$

where n is the total number of courses within the curriculum.

The complexity of the curricula for the SOE and the ASM are shown in Tables 2 and 3 respectively.

The results show that the most complex program in the SOE at UNM is mechanical engineering with a complexity score of 461, whereas the most complex program in the ASM is interdisciplinary film and digital media program (IFDM)

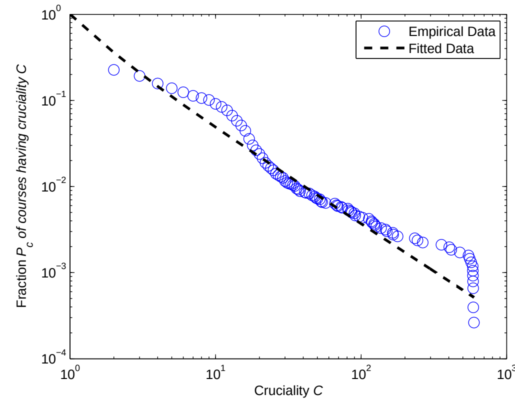


Figure 3: Cumulative distribution function P_c for the cruciality of UNM courses distributed according to power law with alpha value of 2.1.

Course	C_i
ISM 100: Algebraic Problem Solving	596
MATH 101: Intermediate Algebra Part 1	590
MATH 102: Intermediate Algebra Part 2	589
MATH 120: Intermediate Algebra	589
MATH 103: Intermediate Algebra Part 3	588
MATH 121: College Algebra	587
MATH 118: Algebra	586
MATH 150: Pre-Calculus Mathematics	570
MATH 123: Trigonometry	557
MATH 162: Calculus I	547

Table 1: Top 10 crucial courses at UNM.

[1]All the UNM data used in this work are found at s3.amazonaws.com/network-analysis-for-universty-courses/unm-data.zip

Program	*Complexity S*
Mechanical Engineering	461
Chemical Engineering	440
Electrical Engineering	368
Computer Engineering	349
Nuclear Engineering	318
Civil Engineering	294
Construction Engineering	271
Construction Management	215
Computer Science	197

Table 2: Complexity of SOE curricula at UNM.

Program	*Complexity S*
IFDM	247
Finance	208
Marketing Management	189
Operations Management	186
Organizational Leadership	177
Human Resource Management	173
Accounting	172
Entrepreneurship	171
General Management	169
International Management	165
MIS	146

Table 3: Complexity of ASM curricula at UNM.

with a complexity score of 247. On the other hand, the least complex program in the SOE is computer science, and management information systems (MIS) is the least complex program in the ASM. Based on these results we were able to compare the average complexity of the curricula in the SOE with that of the ASM. Table 4 shows that curricula in the SOE are, on average, more complex than those in the ASM. This result might be predictable especially for the fact that analyses done on UNM students data for the years 2006 and 2007 show that the percentage of engineering students who finished their programs within the specified eight semesters is much lower than UNM business students. This is shown in Table 4. Note again that the difficulty level of the courses is not taken into consideration in this work.

College	*Average Complexity*	$\leq$ 8 semesters (2006-07)
SOE	324	8.3 %
ASM	182	28.7 %

Table 4: Average complexity of the curricula in the SOE and the ASM at UNM

Our framework may also be used to detect the most crucial courses in a given college. Hence, this would allow academic and faculty leaders to provide more resources and efforts to such courses, which may positively impact the academic progress of students within these colleges. We applied this detection algorithm to the courses in SOE and ASM programs at UNM. For the SOE, we merged all the curricula into one graph. As previously mentioned, cycles are eliminated and transitive edges are deleted. We applied the same procedure to curricula in the ASM. The resulting top 10 crucial courses in the SOE and the ASM are shown in Table 5.

Algorithm 1 Degree Program Algorithm.

Define Cr to be the list of courses within a curriculum
Define M to be the adjacency matrix representation of the curriculum
Define T to be the number of terms in the curriculum
Define N to be the initial number of courses in each term: $N = \lfloor \frac{\text{Length of } Cr}{T} \rfloor$
Initialize t =1
while Cr not empty and $t \leq T$ **do**
  Calculate C_i for all the courses in Cr
  Choose N course (or less in case the remaining courses for the last term are less than N) from Cr with the highest C_i and that do not have any pre-requisite(s)
  The chosen courses will fill in term t
  Remove the chosen courses from Cr and M
  $t = t + 1$
end while
Initialize t' =1
while Cr is not empty **do**
  Calculate C_i for all the remaining courses in Cr
  Choose the course with the highest C_i
  The chosen course will be added to courses in term t'
  Remove the chosen course from Cr and M
  $t' = t' + 1$
end while

5. APPLICATIONS

5.1 Creating Optimal Curriculum

Based on the previous analyses, the cruciality framework may also be used to construct a optimal curricula for degree programs. In other words, given specific courses, the number of terms (semesters) for a degree program, and the pre-requisite dependency relationships, it is easy to build a curriculum that provides optimal choices for the courses to be taken in each term. Thus the risk of delay for a student is lowered by following the suggested curriculum. The Degree Program algorithm is detailed in Algorithm 1.

Besides generating an optimal curriculum, the Degree Program algorithm may also be expanded to generate different curriculum versions based on students' adjustments. In other words, a student might want, for various reasons, to take different courses than those suggested by the optimal curriculum in term 1 for example. Then, based on the student choices for courses in term 1, the algorithm would suggest courses to be taken for the remaining terms. In this case, the algorithm might alert students that they might end up heavyly loaded for the remaining terms, or that their graduation may even be delayed.

5.2 Student Progress

As mentioned in previous sections, time-to-degree is a critical factor in the academic life of both students and universities. Students normally want to obtain their degrees as soon as possible (subject to financial and work-life constraints) while universities want their graduation rate to be as high as possible. Usually grades (e.g., GPA) are the main criteria to measure the students' progress throughout a curriculum. Grades do not however take the time factor into consideration. Theoretically, a student may have a high GPA while progressing slowly through the curriculum. Engineering student A who took crucial courses in the first semester earning

SOE		ASM	
Course	C_i	*Course*	C_i
MATH 162: Calculus I	115	ENGL 101: Composition I: Exposition	35
MATH 163: Calculus II	79	ENGL 102: Composition II: Analysis	33
MATH 121: College Algebra	73	MATH 121: College Algebra	23
ECE 131: Programming Fundamentals	73	ECON 106: Introductory Microeconomics	23
PHYC 160: General Physics	61	STAT 145: Introduction to Statistics	18
CS 151L: Computer Programming Fundamentals	60	MGMT 202: Principles of Financial Accounting	17
MATH 316: Ordinary Differential Equations	55	MGMT 322: Marketing Management	14
CHEM 121: General Chemistry I	49	MATH 180: Elements of Calculus I	13
CHEM 123L: General Chemistry I Lab	48	IFDM 105L: Inter and New Media Studies I	13
CHEM 122: General Chemistry II	39	MGMT 306: Organizational Behavior and Diversity	12

Table 5: Top 10 crucial courses in the SOE and ASM programs at UNM.

a GPA of 4.0 is therefore in better shape than student B who took non-crucial courses while earning the same GPA of 4.0. Obviously, the probability that student B may be delayed in a program is higher than that of student A based on the definition of crucial courses. Hence and based on the time factor mentioned, crucial courses must be included in studying the progress of students through out their respective academic life.

5.2.1 Student Progress Framework

To achieve this, we propose a framework that makes use of the Degree Program algorithm, the cruciality parameter, and the earned letter grade. We thus create an optimal curriculum for every department within the university and accordingly monitor a student's progress every semester based upon the type of the courses (i.e., crucial or noncrucial) taken, along with respective letter grades. Students having more courses matching the cruciality values of the optimal curriculum courses per term are in a better shape assuming all students have the same GPA. Figure 4 shows a 3-term optimal curriculum. Next to each course there are two numbers. The number on top represents the cruciality value whereas the one below represents the earned letter grade. Note that in this work the highest grade value of a course is 4.0. Assuming students X and Y have the same letter grades for all the courses as shown in Fig. 4, student X is less likely to get delayed throughout her academic program. Numerically, this may be quantified by summing the product of both cruciality value and letter grade of all the courses taken up to that term, that is:

$$P_j = \frac{\sum_{ij} C'_{ij} G_i}{\sum_{in} C'_{in}} \quad (4)$$

where P_j is the student progress score (SPS) at term j, C'_{ij} is the cruciality value of course i taken in term j, G_i is the letter grade of course i and n is the total number of terms in a curriculum. Note that a course must match one of the optimal curriculum courses otherwise the cruciality value is zero, that is

$$C'_{ij} = \max\{C_i, 0\} \quad (5)$$

The denominator in Eq. (4) normalizes the SPS so that P_j is always less than or equal to 4.0. In fact the value of SPS in the last semester is equivalent to the GPA value. However, the advantage of SPS over GPA is its ability to quantify student progress, taking into consideration the time factor mentioned previously. Its criticality is even more evident in the first couple of semesters where students at this time need more advisement than at other points in their academic careers. Examples shown in the next section illustrate how to analyze SPS in order to provide advisement when needed.

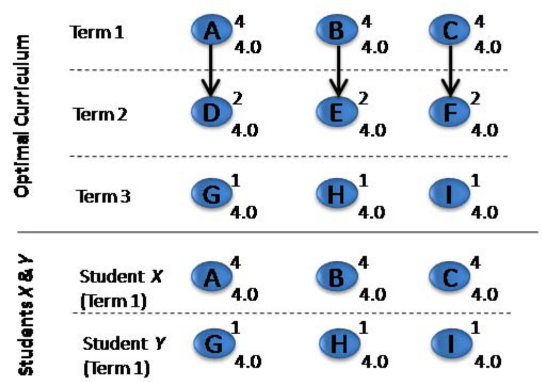


Figure 4: Progress of students X and Y with respect to the optimal curriculum. SPS of X is $\frac{48}{18}$ whereas that of Y is $\frac{12}{18}$.

As a first step, the student progress framework discussed above is to an extent idealistic. Normally, curricula are not as simple as it appears in Fig. 4. For example, degree requirements for many curricula are technical elective courses, social science courses, humanity courses, etc. It may therefore be hard to create one optimal curriculum for a particular program. Some technical elective or social science courses might have different cruciality value. Thus SPS would not reflect the true progress value unless some further assumptions are made. First it should be clear that there must be one curriculum for each program. Accordingly, this means there must be one reference to which students can refer to and hence student progress framework would be feasible then to apply. To achieve this, it is assumed that all degree requirements that are unspecified within a curriculum (e.g. technical elective courses and social science courses) do not have pre-requisites. This will provide a curriculum with a minimum bound above which the SPS wouldn't exceed. So all the unnamed courses taken by a student in which they match her respective department curriculum are assumed to have no pre-requisites and thus the cruciality values for such courses are 1. Note that the cruciality values of the courses in each term in the optimal curriculum are computed excluding the pre-requisite edges emerging from courses of previous terms. Hence, for example, the cruciality values of the courses D, E and F in Fig. 4 are 1 instead of 2. Thus, once students pass courses A, B and C, it makes no difference if they take D, E and F before G, H and I or vice-versa the next term. The example shown in Fig. 5 illustrates the main idea of the student progress framework. The optimal SPS P_1^o, P_2^o and P_3^o are $\frac{48}{17}$, $\frac{60}{17}$ and $\frac{68}{17}$ respectively.

5.2.2 *Student Progress Ratio*

Analyzing student progress is attained by considering the ratio of P_j^s from the student curriculum over that of P_j^o from the optimal one each term, that is:

$$I_j = \frac{P_j^s}{P_j^o} \tag{6}$$

where I_j is the student progress ratio (SPR) at term j. If the value of SPR is greater than or equal to 1, then the student is on track. Otherwise, special attention must be taken depending on far below 1 the SPR is. For example, in Fig. 5, $I_1 = \frac{24}{48}$ which is 0.5. This might be a sign that student X is at risk of being delayed, because she did not take crucial courses in her second term or because she earned bad grades. Note that SPR must be less than or equal to 1 in the last term meaning that the student has finished all the curriculum requirements.

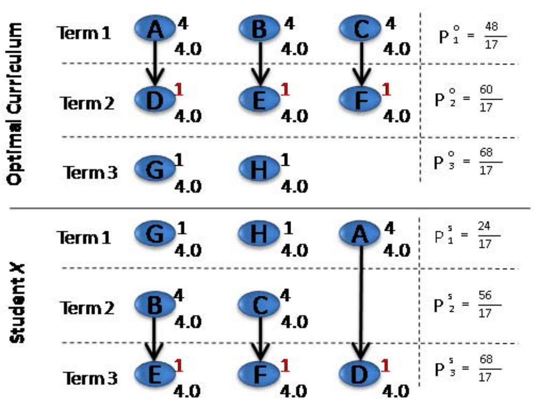


Figure 5: SPR of student X. $I_1 = \frac{24}{48}$; $I_2 = \frac{56}{60}$; $I_3 = \frac{68}{68}$.

6. CONCLUSION

Using complex network analysis and graph theory, we proposed a framework to study the structure of university course networks. We introduced a measure of the cruciality of courses at three different levels: the university, a college and a department. Based on real data, our results show that the courses at UNM follow a power law distribution. The results also show that our proposed framework is suitable to study the complexity of a curriculum within a college. Based on longest path and connectivity, it was shown that the curriculum in the SOE at UNM is, on average, more complex than that in the ASM. The framework is extended further to create an optimal curriculum for a department where optimality is characterized by lessening the risk of delayed graduation. We also introduced the SPR metric to study the progress of a student in a curriculum, and hence predict progress based on the grades of the courses taken, the number of courses, as well as their respective cruciality values. This application is therefore very useful in tracking the progress of students and to intervene (via advisement, academic support, etc.) in order to have a positive impact on graduation rates.

7. REFERENCES

[1] Educational attainment in the united states. United States Census Bureau, 2012.

[2] Performance funding for higher education. National Conference of State Legislatures, Feb. 2012.

[3] A. V. Aho, M. R. Garey, and J. D. Ullman. The transitive reduction of a directed graph. *Society for Industrial and Applied Mathematics*, 1(2), June 1972.

[4] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, Nov. 2009.

[5] L. DeAngelo, R. Franke, S. Hurtado, J. Pryor, and S. Tran. Completing college: Assessing graduation rates at four year colleges. Technical report, Higher Education Research Institute, UCLA, Los Angeles, CA, Nov. 2011.

[6] L. K. Lau. Institutional factors affecting student retention. *General Education Journals*, 124(1), 2003.

[7] A. Venezia, P. M. Callan, J. E. Finney, M. W. Kirst, and M. D. Usdan. The governance divide: A report on a four-state study on improving college readiness and success. Technical report, The National Center for Public Policy and Higher Education, San Jose, CA, Sep. 2005.

[8] L. Zhang. Does state funding affect graduation rates at public four-year colleges and universities? *Educational Policy*, 23(5):714–731, Sep. 2009.