Journal of Quality Measurement and Analysis *JQMA* 20(2) 2024, 127-147
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2002.2024.10

LEARNING ANALYTIC FRAMEWORK FOR STUDENTS' ACADEMIC PERFORMANCE AND CRITICAL LEARNING PATHWAYS

(Rangka Kerja Analitik Pembelajaran untuk Prestasi Akademik Pelajar dan Laluan Pembelajaran Kritikal)

JESSICA TAN YEN LYN*, GOH YONG KHENG, LAI AN CHOW & NGEOW YOKE MENG

ABSTRACT

In the domain of higher education, the need to leverage data-driven insights for understanding and enhancing student academic performance is becoming increasingly critical. To address this, a unified learning analytics framework is proposed, aimed at deciphering complex student academic journeys and fostering data-informed decision-making for educational institutions. This framework's methodology involves several key steps, starting with standardized data collection and pre-processing. Subsequently, dimensionality reduction techniques like Principal Component Analysis (PCA) and Non-negative matrix factorization (NMF) are applied to capture the most influential course components and grade information. The resulting reduced dataset is then subjected to various clustering algorithms, including partition-based clustering (K-means), hierarchical clustering, and density-based clustering (DBSCAN). These algorithms group students based on academic performance and course profiles, facilitating the identification of clusters with similar characteristics and academic trajectories. Furthermore, a collective network graph is constructed to analyze course relationships and program pathways, identify critical courses, and reveal influential factors affecting student performance and outcomes. This network analysis enables educators to identify bottleneck courses and areas that may require additional support or improvement, fostering a data-driven approach to curriculum design and enhancement. To showcase the framework's efficacy, a case study was conducted on 3550 undergraduates from an engineering program at a Malaysian private university. The student dataset used in this study spans from 2005 to 2021, covering a wide range of academic years for analysis. The results demonstrate the framework's capability to unveil valuable insights into students' academic journeys, revealing key factors contributing to their success. By providing a holistic perspective of student performance and course interactions, the proposed learning analytics framework holds great promise for educational institutions seeking data-driven strategies to enhance student outcomes and optimize learning experiences.

Keywords: learning analytic framework; Principal Component Analysis (PCA); Non-negative Matrix Factorization (NMF); clustering; network graph

ABSTRAK

Dalam domain pendidikan tinggi, keperluan untuk memanfaatkan cerapan dipacu data untuk memahami dan meningkatkan prestasi akademik pelajar menjadi semakin kritikal. Untuk menangani perkara ini, rangka kerja analisis pembelajaran bersatu dicadangkan, bertujuan untuk mentafsir perjalanan akademik pelajar yang kompleks dan memupuk pembuatan keputusan berdasarkan data untuk institusi pendidikan. Metodologi rangka kerja ini melibatkan beberapa langkah utama, bermula dengan pengumpulan data dan pra-pemprosesan piawai. Selepas itu, teknik pengurangan dimensi seperti Principal Component Analysis (PCA) dan Non-negative Matrix Factorization (NMF) digunakan untuk menangkap komponen kursus dan maklumat gred yang paling berpengaruh. Dataset terkurang yang terhasil kemudiannya tertakluk kepada pelbagai algoritma pengelompokan, termasuk pengelompokan berasaskan partition (K-means), pengelompokan hierarki dan pengelompokan berasaskan ketumpatan (DBSCAN). Algoritma ini mengumpulkan pelajar berdasarkan prestasi akademik dan profil kursus, memudahkan pengenalpastian kluster dengan ciri dan trajektori akademik yang serupa. Tambahan pula, graf rangkaian kolektif dibina untuk menganalisis hubungan kursus dan laluan program, mengenal

pasti kursus kritikal dan mendedahkan faktor-faktor yang mempengaruhi prestasi dan hasil pelajar. Analisis rangkaian ini membolehkan pendidik mengenal pasti kursus kesesakan dan bidang yang mungkin memerlukan sokongan atau penambahbaikan tambahan, memupuk pendekatan terdorong data untuk reka bentuk dan peningkatan kurikulum. Untuk mempamerkan keberkesanan rangka kerja tersebut, satu kajian kes telah dijalankan ke atas 3550 mahasiswa dari program kejuruteraan di universiti swasta Malaysia. Hasilnya menunjukkan keupayaan rangka kerja untuk mendedahkan pandangan berharga tentang perjalanan akademik pelajar, mendedahkan faktor utama yang menyumbang kepada kejayaan mereka. Dengan menyediakan perspektif holistik prestasi pelajar dan interaksi kursus, rangka kerja analitik pembelajaran yang dicadangkan memegang janji besar untuk institusi pendidikan yang mencari strategi dipacu data untuk meningkatkan hasil pelajar dan mengoptimumkan pengalaman pembelajaran.

Kata kunci: rangka kerja analisis pembelajaran; Principal Component Analysis (PCA); Non-negative Matrix Factorization (NMF); pengelompokan; graf rangkaian

1. Introduction

In the realm of educational research and pedagogy, the advent of Educational Data Mining (EDM) and Learning Analytics has opened up exciting possibilities for enhancing the learning experience and academic performance of students. As data-driven approaches gain momentum in education, researchers have explored various methodologies to analyze educational datasets and uncover valuable insights. Despite the vast amount of data available in educational systems, there is a significant challenge in effectively utilizing this data to understand students' academic performance and identify critical learning pathways. The absence of a comprehensive learning analytic framework limits educational institutions' ability to gain insights into the factors influencing student success and design targeted interventions (Khalil *et al.* 2022). Existing approaches often focus on individual aspects of data analysis, such as academic performance or course components, without considering the holistic view of students' learning journeys (Omar *et al.* 2020; Su *et al.* 2023). This fragmented approach hinders the ability to identify influential courses, trace critical learning pathways, and provide personalized support to students. Therefore, there is a pressing need for a cohesive learning analytic framework that incorporates educational data mining techniques to analyze students' academic performance, cluster course components, and trace critical learning pathways. Addressing this problem will enable institutions to make data-driven decisions, enhance instructional practices, and improve student outcomes in a more comprehensive and meaningful manner.

The primary goal of this research is to construct a unified Learning Analytic Framework that harnesses the potential of clustering algorithms and network analysis to gain a comprehensive understanding of student's academic performance and critical learning pathways. By clustering students based on their learning behaviors and applying network analysis to explore the patterns of interactions and knowledge exchange, this framework seeks to identify significant relationships between academic performance, course components, and learning strategies. By investigating the potential synergies between these methodologies, this research seeks to contribute to the development of data-driven, evidence-based educational practices that can significantly improve the learning experience and academic outcomes of students.

2. Related Works

Considering the collective insights from these articles, constructing a unified learning analytic framework for this study can prove highly beneficial. By integrating clustering, network analysis, and other data mining techniques, this unified framework can serve as a powerful tool for educators to gain actionable insights into students' academic performance and design targeted

interventions to address their learning challenges effectively. Ultimately, this approach aligns with the study goal of leveraging data-driven methodologies to improve teaching practices, enhance student engagement, and boost overall learning outcomes.

2.1. *Learning analytic framework*

There are relevant articles that propose learning analytic frameworks to analyze and measure various aspects of students' performance and success in educational settings. Bharara *et al.* (2018) presents an application of learning analytic using clustering data mining to analyze students' dispositions. Almond-Dannenbring *et al.* (2022) present a comprehensive framework for student success analytic which involves the integration of various data sources, such as academic records, learning management system data, and student engagement data. Joshi *et al.* (2020) proposed a learning analytic framework tailored for engineering education with a focus on problem-based learning, aimed to measure students' performance and teachers' involvement during problem-based learning activities. These studies have demonstrated the significance of learning analytic frameworks in educational settings as these frameworks offer valuable insights to educators and institutions to enhance teaching practices, foster student development, and improve overall learning outcomes (Bharara *et al.* 2018; Almond-Dannenbring *et al.* 2022; Joshi *et al.* 2020).

2.2. *Educational data mining*

Recent research has explored the use of clustering analysis in the field of educational data mining, with a particular emphasis on students' performance prediction and academic analysis. Abu Saa (2016) explores the application of educational data mining for students' performance prediction. By identifying distinct clusters of students, the research aimed to predict their performance and learning outcomes. Dol and Jawandhiya (2023) present a comprehensive survey on the combination of classification techniques with clustering and association rule mining in educational data mining. The article highlighted how clustering analysis can be effectively integrated with classification algorithms to improve the accuracy of performance prediction models and uncover hidden patterns and associations in educational datasets. Križanić (2020) present a case study on educational data mining, utilizing both cluster analysis and decision tree techniques. The study applies clustering to group students based on various characteristics and then employs decision trees to understand the factors influencing students' academic performance. This combination of techniques allows for a more nuanced analysis of the data and provides educators with deeper insights into the key determinants of student success. Govindasamy and Thambusamy (2018) focus on analyzing student academic performance using clustering techniques and identifying distinct groups of students based on their performance patterns. These studies have shown that clustering analysis in educational data mining can be used to identify groups of students with similar learning behaviors, learning styles, or learning difficulties and gain insights into factors influencing academic outcomes. The combination of clustering with other data mining techniques enhances the understanding of student behavior and performance, facilitating data-driven decision-making in educational settings (Abu Saa 2016; Dol & Jawandhiya 2023; Križanić 2020; Govindasamy & Thambusamy 2018).

2.3. *Network analysis*

Meanwhile, the significance of network analysis, as seen in the article on social network analysis in higher education online learning by Jan *et al.* (2019), reinforces the idea of using network analysis to identify distinct clusters of students based on their interactions and relationships, further supporting the development of a comprehensive learning analytic framework. Furthermore, a study by Saqr *et al.* (2018) shows that leveraging network visualizations was able to guide educational interventions as the author provided a practical application of using network analysis

for monitoring and improving collaborative learning through enhancing student interactions.

Network analysis has proven to be a valuable and promising approach in the field of education, although there is still a lack of extensive research in this area. Several studies collectively showcase the utility of network analysis in education and have highlighted the usefulness of social network analysis (SNA) to gain insights into educational contexts and foster informed interventions (Samanta *et al.* 2021; Saqr *et al.* 2018; Grunspan *et al.* 2014).

As outlined by Borgatti and Everett (2006), centrality measures such as degree, closeness, and betweenness can provide insights into the importance of nodes within a network. While degree centrality is based only on the number of connections, closeness, and betweenness centralities incorporate the broader network structure to identify nodes that facilitate communication over short paths or bridge connections between others Borgatti and Everett (2006). Hence, this study identified the most influential courses from the network analysis of students from different programmes.

3. Methodology

The methodology employed in this study begins with data collection, where the input dataset format of the learning analytic framework is designed to follow a specified format that is common to all institutions. This ensures consistency and allows for seamless integration and analysis of student data from different sources. After data collection, the next step is data cleansing, which involves removing missing data and unknown grade entries. This includes cases such as exempted courses, withdrawal students, credit transfers, and other similar scenarios that may result in incomplete or unreliable data. Once the data is cleansed, the grade information is transformed into a Grade Point Average (GPA) according to institution standards. The study then proceeds to calculate the weight matrix for each student. This is achieved by multiplying the GPA of each course with the course-relation matrix. The course-relation matrix indicates the order in which courses are taken based on the students' session intakes, providing insights into the sequencing of courses within the curriculum. To reduce the dimension of the weight matrix and capture the essential information, dimensionality reduction techniques were applied. This step helps in managing the complexity of the data and extracting meaningful patterns from the weight matrix, allowing for more efficient analysis and interpretation. Furthermore, the students are clustered based on their weight matrices. Clustering algorithms are employed to group students with similar weight matrices, enabling the identification of distinct clusters based on academic performance and course sequencing. This clustering process aids in understanding the patterns and similarities among students, facilitating further analysis of critical learning pathways and performance trends. Finally, a network graph was constructed to identify the critical courses and the interconnections between courses and allow for the identification of both common and critical learning pathways. Figure 1 shows the proposed methodology of this study.

Figure 1: Overview of proposed learning analytic framework

3.1. *Data collection*

To ensure a common input dataset format that can be applied to student datasets from multiple institutions, the following information should be included:

(1) Student ID: A unique identifier assigned to each student in the dataset. This allows for individual student tracking and analysis.
(2) Session: The specific intake or enrollment period in which the student joined the institution. This helps capture the temporal aspect of the data and enables the analysis of academic performance over time.
(3) Course Units: The name or code of the course taken by the student. This provides information about the specific courses in which students are enrolled.
(4) Total Mark: The total marks achieved by the student in a particular course. This indicates the student's performance in terms of numerical scores or percentages.
(5) Grade: The grade obtained by the student in the course. This can be represented as letter grades (e.g., A, B, C) or a grading scale specific to the institution.
(6) Programme: The program or field of study pursued by the student. This helps in analyzing academic performance within specific programs or disciplines.

Table 1: Sample of the partial input dataset

Student_Id	Session	Course_Unit	Total_Mark	Grade	Programme
1	200705	Math1	63	B-	Engineering1
1	200805	Core1	#	#	Engineering1
1	200805	Core4	51	C	Engineering1
3	200810	Core4	79	A-	Engineering1
3	200805	Core1	54	C	Engineering1
4	200810	Arts1	96	A	Engineering1
4	200810	Math2	78	A-	Engineering1
4	200810	Math3	#	#	Engineering1

Table 2: Grade-to-GPA Mapping

Grade	GPA
A+	4.00
A	4.00
A-	3.67
B+	3.33
B	3.00
B-	2.67
C+	2.33
C	2.00
D	1.00
F	1.00

3.2. *Data pre-processing*

Data pre-processing for a student dataset typically involves cleaning and preparing the data for analysis. In this study, the data cleansing stage includes removing unknown grading entries and eliminating empty or incomplete data points. Any entries that represent unknown, invalid, or incomplete values will be excluded.

Next, the data transformation stage involves GPA conversion which is a critical step in the pre-processing of educational datasets, especially when dealing with diverse grading systems. To ensure uniformity and consistency, a mapping or lookup table is typically created to associate each unique grade with its corresponding GPA value. Table 2 acts as a reference guide for the system to translate various grades into a standardized Grade Point Average.

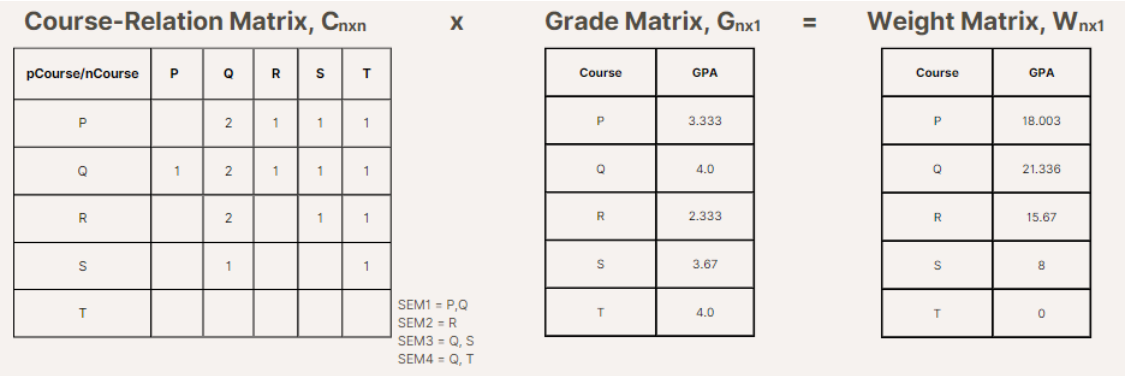
3.3. *Weight matrix*

This study utilized a weight matrix as the input data for clustering analysis. The weight matrix represents the relationship between students' grades and their course sequences. It was derived from the product of the grade matrix and the course relation matrix for each student, see Figure 2 for example.

Steps to compute weight matrix:

(1) Create an $n \times n$ course-relation matrix, $\boldsymbol{C_{n \times n}}$ that tells which course comes first and which course is taken next based on the student dataset.
(2) Create an $n \times 1$ grade matrix, $\boldsymbol{G_{n \times 1}}$ that tells the student achievement in each course in terms of GPA.
(3) Compute the weight matrix, $\boldsymbol{W_{n \times 1}}$ for each student by performing dot product for the course-relation matrix and grade matrix. Each student was represented by a unique

weight matrix, capturing their academic performance and course relationships.



pCourse/nCourse	P	Q	R	S	T
P		2	1	1	1
Q	1	2	1	1	1
R		2		1	1
S		1			1
T					

Course	GPA
P	3.333
Q	4.0
R	2.333
S	3.67
T	4.0

Course	GPA
P	18.003
Q	21.336
R	15.67
S	8
T	0

Figure 2: Example of weight matrix computation

Besides, to examine the impact of different grading scenarios on clustering results, this study explored three distinct test cases: the worst-case grade, best-case grade, and average-case grade achievement. These test cases reflect variations in the weight matrix, which is influenced by the product of the grade matrix and the course relation matrix for each student. Hence, each test case offered unique insights into student clustering, taking into account different grading approaches and variations in course retakes.

The worst-case grade test case focused on analyzing the weight matrix by retaining only the lowest GPA achieved in each course for each student. This scenario becomes relevant when students have retaken courses multiple times, resulting in multiple GPA records for the same course unit. By retaining the worst grade, this study aimed to understand the clustering patterns for students who faced challenges or struggled academically. This test case allowed us to identify clusters representing students who may have encountered difficulties in their learning journey, such as those who needed additional support or experienced setbacks. Conversely, the best case grade test case considered only the highest GPA among all the achievements in each course for each student. This scenario aimed to capture the clustering patterns of academically high-performing students who consistently excelled in their coursework. By selecting the best grades attained, we aimed to identify clusters representing top-performing students who consistently achieved exceptional academic results. This test case shed light on the characteristics and attributes of high-achieving students within the identified clusters. Additionally, we explored the average grade test case. This test case involved calculating the average GPA across multiple retakes for the same course unit. It aimed to assess the clustering patterns for students whose academic performance varied across multiple attempts at the same course. By considering the average GPA, we aimed to identify clusters representing students with relatively consistent or average academic performance. This test case provided insights into clusters comprising students who may have experienced both successes and challenges in their academic journey.

It is important to note that these test cases were particularly meaningful when students had retaken courses multiple times and when there were multiple GPA records for the same course unit. The variations in the weight matrix among these test cases reflect different grading approaches and the influence of course retakes or multiple attempts on academic performance. By considering different weight matrices, we could assess the impact of these variations on the clustering outcomes and gain insights into the characteristics and academic trajectories within each cluster. The clustering analysis using the weight matrix allows us to identify clusters of students who share similar academic profiles and potentially uncover patterns related to their learning experiences.

3.4. *Dimensionality reduction*

This learning analytic framework explored the impact of different dimensionality reduction methods on the clustering results, alongside variations in the weight matrix based on different grading scenarios. Multiple dimensionality reduction techniques, including PCA, and NMF, were applied to process the weight matrix and obtain lower-dimensional representations. The selection of various dimensionality reduction methods allowed for comprehensive coverage of different input student datasets.

PCA proved effective for weight matrices exhibiting linear relationships, effectively preserving the overall data structure and variability (Greenacre *et al.* 2022). NMF was particularly suitable for the non-negative nature of GPA achievements, representing the data as a combination of non-negative basis vectors and coefficients (Leuschner *et al.* 2019).

The optimal number of components for each method was determined through the elbow method, a data-driven approach that identifies a significant drop in explained variance or reconstruction error. By automatically applying the elbow method, this study was able to select the optimal number of components for each dimensionality reduction method. This automated approach ensured that we retained an appropriate number of components that captured the most essential information while avoiding overfitting or underfitting the data. It enabled us to strike a balance between dimensionality reduction and preserving the key characteristics of the weight matrix.

Each dimensionality reduction technique offered unique insights into the clustering structure and patterns, allowing for the identification of the most suitable algorithm for each dataset based on clustering scores. The automation of this process enhanced the reliability and reproducibility of the results, mitigating any potential subjective biases that could arise from manual component selection for dimensionality reduction. As a result, the framework demonstrated robustness and interpretability in analyzing student academic performance and course patterns, contributing to its overall effectiveness in the research study.

3.5. *Clustering methods*

In the clustering stage, the framework performed clustering on students' weight matrices obtained from different dimensionality reduction methods for each test case. Various unsupervised learning algorithms like K-Means, DBSCAN, and hierarchical clustering were used to group students with similar academic performance and course patterns, facilitating the identification of distinct clusters. Jain *et al.* (1999) has discussed the principle of different clustering algorithms where DBSCAN, a density-based algorithm, identifies dense regions as clusters and handles noisy data effectively. On the other hand, K-Means partitions data into 'k' clusters, while hierarchical clustering builds a dendrogram hierarchy of clusters.

In performance metrics evaluation wise, Alhajj *et al.* (2022) highlights two widely used metrics in clustering analysis namely the Silhouette score and the Calinski-Harabasz (CH) score. This framework combines these two metrics to select the best-performing clustering algorithm for each test case.

Referring to Shahapure and Nicholas (2020), the Silhouette score measures the cohesion and separation of clusters, providing a measure of how well-defined and compact the clusters are. It takes into account the average distance between data points within clusters and the average distance between data points of different clusters. A higher Silhouette score indicates well-separated clusters with high intra-cluster similarity, suggesting a better clustering result. The equation for the Silhouette score is described in Eq. 1.

$$s(i) = (b(i) - a(i))/max(a(i), b(i)) \tag{1}$$

where

- $s(i)$ is the Silhouette score for data point i

- $a(i)$ is the within-cluster cohesion of data point i
- $b(i)$ is the between-cluster separation of data point i

Referring to Caliński and Harabasz (1974), the Calinski-Harabasz (CH) score, also known as the Variance Ratio Criterion, quantifies the separation between clusters and the compactness within clusters. It compares the between-cluster variance to the within-cluster variance, providing a measure of the clustering's compactness and separation. A higher Calinski-Harabasz (CH) score suggests more compact and well-separated clusters. The equation for the Calinski-Harabasz (CH) score is described in Eq. 2.

$$CH = (n * sum(b))/(sum(w) * (n - k)) \tag{2}$$

where

- n is the number of data points
- k is the number of clusters
- $sum(b)$ is the sum of the between-cluster dispersion
- $sum(w)$ is the sum of the within-cluster dispersion

Then, the algorithm with the highest scores was selected as the best clustering algorithm which might vary across different test cases.

3.6. *Network analysis*

This study conducted further analysis by incorporating network analysis, which involved constructing a network graph based on the course sequences of students within the same cluster. The goal was to examine the relationships and patterns within the course sequences and identify common courses and recurring sequences among the students. Each student's course sequence was represented as nodes and connections or edges were established based on the sequential order of the courses. The network graph visually represented the interconnections between courses.

By analyzing the network graph, common courses were identified through the presence of multiple edges connecting the same pair of nodes or nodes with a higher degree (indicating more connections). The framework used visual cues to highlight differences in node size and edge thickness in the graph. Larger nodes represented courses commonly taken by students within the cluster, while smaller nodes represented less frequently taken courses. Thicker edges indicated a higher occurrence of transitions or more common course sequences within the cluster.

The network analysis helped identify recurring sequences of courses taken by students within the cluster, potentially indicating common learning pathways or program requirements. Centrality measures were also employed to identify critical courses significantly influencing students' academic performance.

The findings from the network analysis complemented the clustering results, providing a more comprehensive understanding of the student's academic profiles and characteristics within each cluster. Integrating network analysis with clustering results offered a holistic view of students' academic performance, course sequences, and common patterns, enhancing insights into their learning trajectories and facilitating data-driven decision-making in educational settings.

4. Results and Findings

In this study, we focused on analyzing the academic performance and critical learning pathways of students from one specific engineering program in a private university in Malaysia. Among the three test cases conducted, we chose to dive deeper into the results of the "First"

test case as an illustrative example. Among the different dimensionality reduction and clustering methods applied, KMeans clustering with NMF dimensionality reduction stood out as the best-performing approach based on the scores obtained for the 'first' test case scenario.

In this test case, the clustering analysis yielded insightful results, uncovering six distinct clusters that provide valuable information about the academic performance and course pathway sequences followed by students as shown in Figure 3 below. Each cluster exhibits unique characteristics, shedding light on the diverse learning patterns and strengths of the student population. The six clusters represent groups of students who share similar academic trajectories and face specific challenges or excel in certain subject areas. The clustering analysis considers both GPA performance in various courses and the sequence of courses taken by students throughout their academic journey.

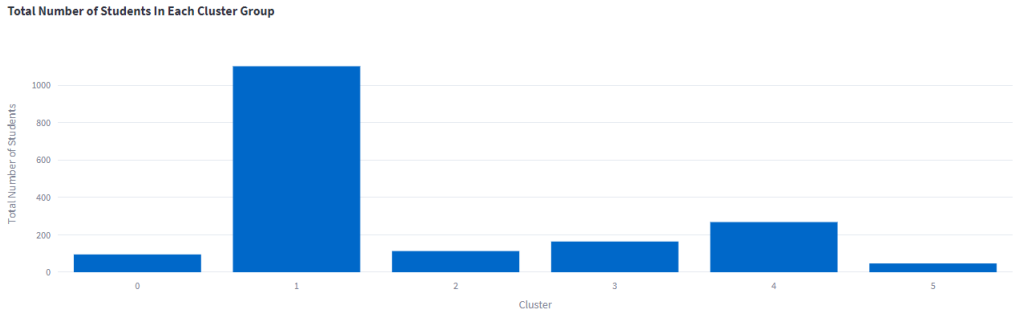


Figure 3: Distribution of Students Across Different Clusters

This learning analytic framework produced unique bar charts for each of the six identified clusters, providing valuable insights into their strengths and challenges across various courses by observing the distinct academic performance of each cluster of students. Each cluster has its unique bar chart, showcasing the average grade achieved by students in different courses. To facilitate a comprehensive understanding, all courses are categorized into five different groups, namely math-related courses (Math), core courses (Core), elective courses (Elective), computer science-related courses (CS), and liberty arts courses (Arts) with each category represented by a distinct color. In the following bar chart in Figure 4 to Figure 13, courses falling under the category of math-related subjects are depicted using a blue color bar. These courses typically involve mathematical concepts and techniques and are essential in engineering programs. Core courses, crucial components of the curriculum, are represented with a light green color bar. These courses are fundamental to the program and often cover core principles and knowledge areas. Elective courses, offering students some degree of choice in their studies, are depicted using a yellow color bar. These courses provide flexibility for students to explore specific interests within the program. Computer science-related courses, which may be part of the curriculum, are represented with a pink color bar. These courses typically cover topics in computer programming and technology. The last category is the liberty arts courses, focusing on humanities and liberal arts, which are depicted using a red color bar. These courses offer students a broader perspective and a well-rounded education.

Besides, the height of each bar corresponds to the average grade attained by students in that particular course. The varying heights of the bars indicate the relative academic performance of students from different clusters in each course. Clusters with higher bars in specific subject categories demonstrate stronger academic proficiency in those areas, while clusters with lower bars may encounter challenges in those courses. This visualization enables a clear comparison of how students from different clusters perform in various courses, pinpointing the areas in which they excel and those where they encounter challenges. Besides, the absence of a bar for a specific course can indicate several possibilities. One of the possibilities might be the students in the cluster have not taken that course due to various reasons, such as the course not being

offered during the period, or it is an elective course that only a few students chose to enroll in. Another possible reason for a missing bar could be that the course has been obsoleted or replaced with a new course code. In educational institutions, courses may undergo updates or changes over time to align with evolving curriculum standards or industry demands. As a result, the old course code might no longer be in use, leading to its absence in the dataset.

Among all the six clusters, Cluster 1 is the largest cluster with 1101 students, comprising a significant portion of the student population in this engineering program. Upon closer examination of the academic performance in Figure 4, students in this cluster show below-average performance in the majority of the courses, including math-related courses (Math), programming-related courses (CS), elective courses (Elective), and core courses (Core). The interconnected nature of these courses means that poor performance in one course may have a cascading effect on others, leading to overall lower academic achievement in this cluster.

Upon analyzing the network analysis graph as shown in Figure 5 below, the red nodes represent the critical courses for Cluster 1 students. These critical courses are primarily mathematics-related. These courses are considered fundamental courses in this engineering curriculum and form the backbone of engineering principles and problem-solving techniques. A concerning finding is that students' performances in these critical courses are consistently below average. Since these courses are interrelated and have a significant impact on other core courses, sub-par performance in critical courses can potentially affect their performance in other correlated courses, leading to negative outcomes in their overall academic journey. Given that Cluster 1 comprises the largest number of students, the majority of students in the engineering program fall into this cluster and follow a similar course sequence. The finding raises an alarm for the institution to assess and potentially revamp the course structure arrangement. Addressing the challenges in these critical courses can have a significant impact on improving the overall academic performance of students in the engineering program.

The network graph also reveals the common pathway that most students in Cluster 1 follow, represented by thicker black edges. Notably, the common sequence involves core courses such as Core8, Core9_b, Core12_b, Core14, and Core16. This indicates that most students are adhering to the program's course structure plan, as deviations from this common pathway are rare. Apart from this, these courses are typically taken during the second year of the engineering program and are designed to provide students with fundamental knowledge and skills in various aspects of specific engineering. Many of these second-year core courses are interrelated and build upon each other. For instance, Core8 (Introductory Electromagnetics) provides the foundation for Core14 (Process Control and Instrumentation), and both courses are relevant to Core12_b (Power Systems).

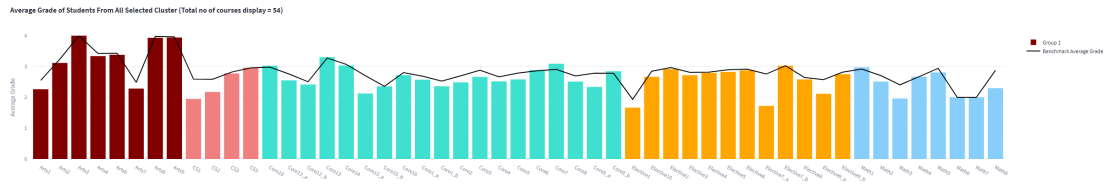


Figure 4: Bar chart depicting the average grades of students in Cluster 1 across various courses

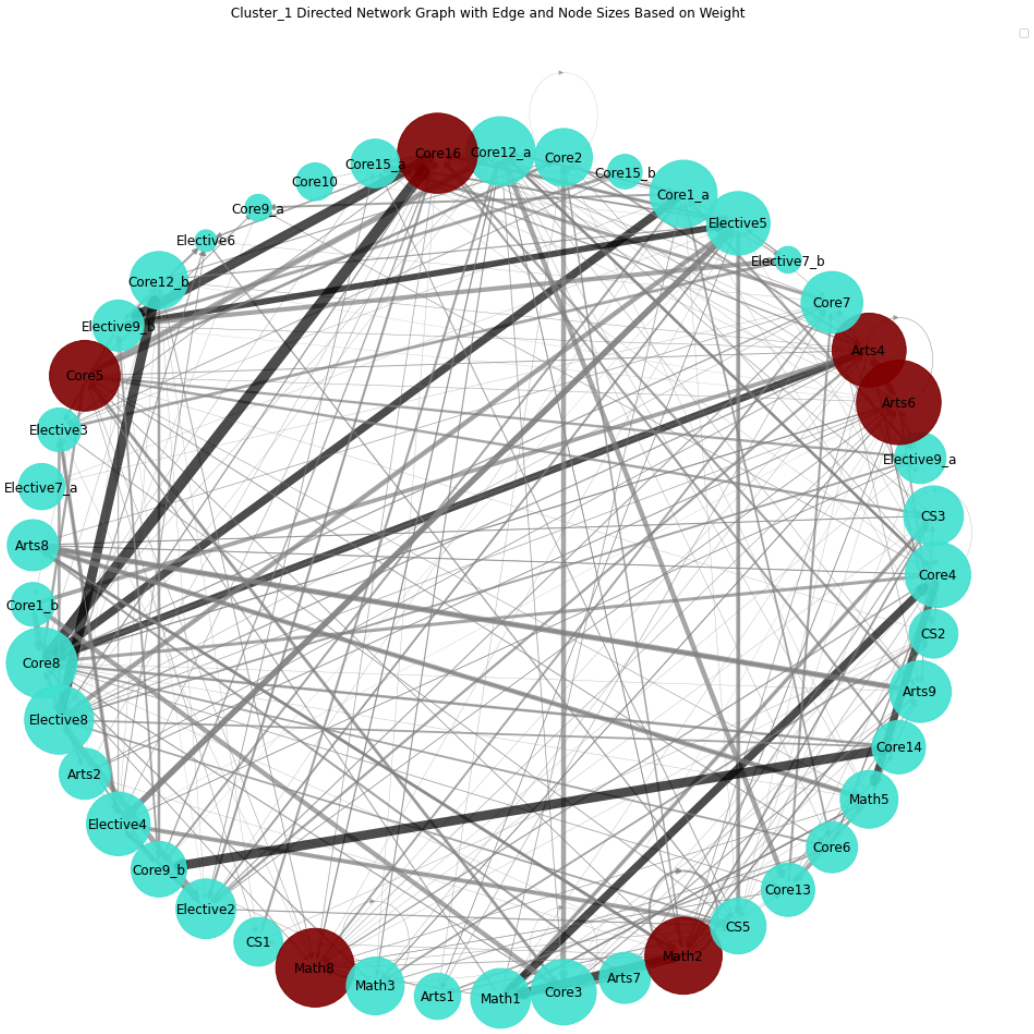


Figure 5: Network graph of Cluster 1

Cluster 3 represents another unique group of students with distinct characteristics in their course selections and academic performance. This cluster was characterized by the absence of computer science-related courses and some mathematics-related courses, indicating a selective preference in their course selection, see Figure 6. In terms of academic performance, students in Cluster 3 show weaker performance in liberty arts and mathematics courses. However, they exhibit better performance in power system-related courses. Notably, the common elective courses chosen by most students in this cluster also align with power system-related subjects, suggesting a particular interest in this area of study.

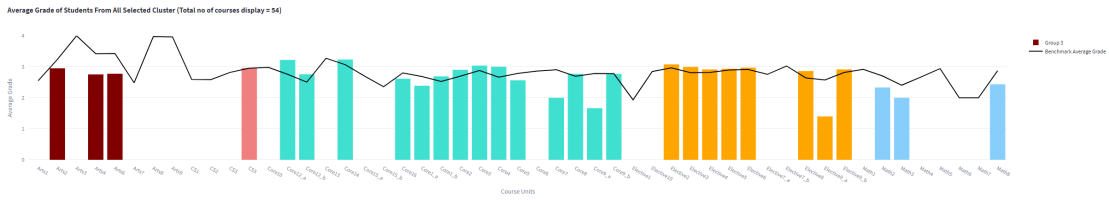


Figure 6: Bar chart depicting the average grades of students in Cluster 3 across various courses

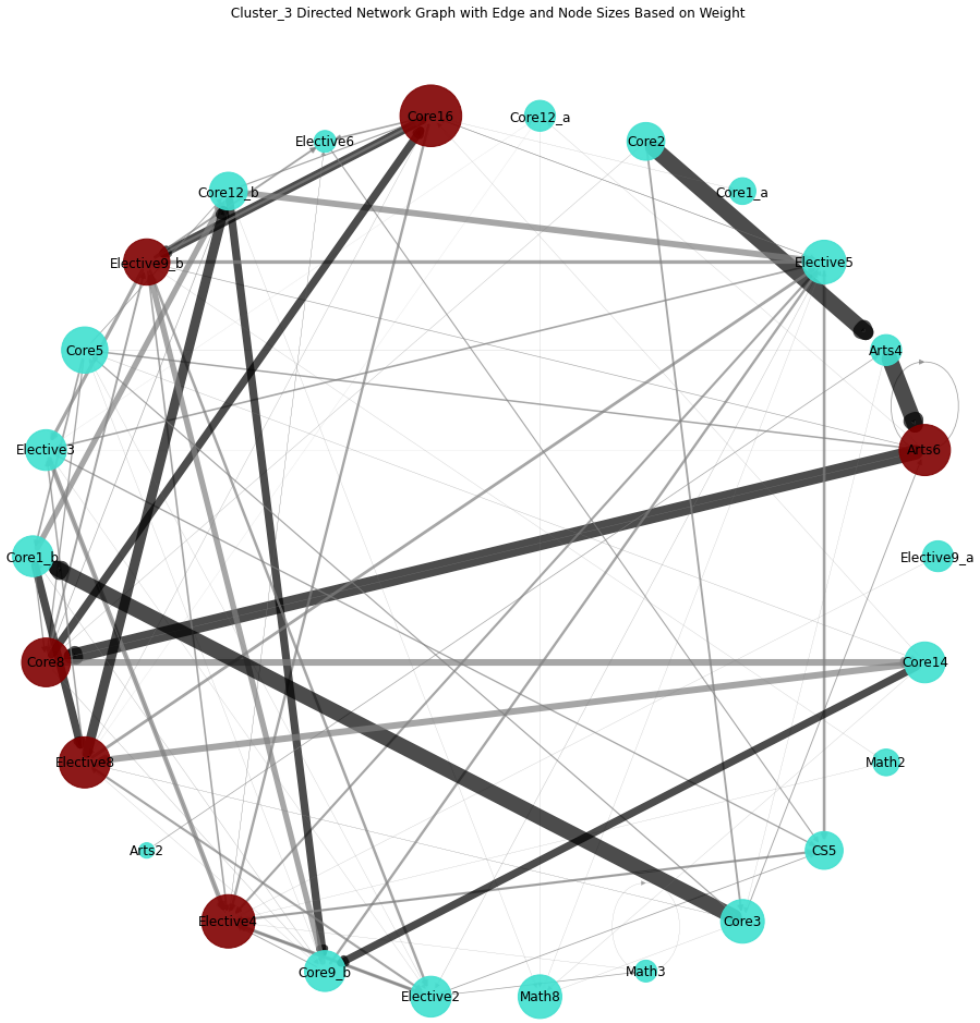


Figure 7: Network graph of Cluster 3

The analysis of the academic years of students in Cluster 3 indicates that they primarily belong to the academic years from 2006 to 2011. This observation is noteworthy, as the course names in this cluster contain the suffix "_b," which suggests that some of these courses have been obsoleted and replaced by "_a" versions in later years. This finding may indicate a time-sensitive aspect of the course structure and could be related to curriculum changes over time. Interestingly, students in Cluster 3 exhibit better performance in power system-related courses, indicating a particular affinity or aptitude for this area of study. The critical courses observed in Figure 7 include power electronics, power systems, and control system technologies, which are essential components of the electrical engineering industry. These critical courses are highly relevant to the electrical industry, as they form the backbone of power electronic systems and control system technologies. These critical courses play a crucial role in shaping the students' expertise and knowledge in these specific areas, which align with their better performance in power system-related courses.

The network graph further provides valuable insights into the common pathways followed by students in this cluster. Two distinct common pathways are identified, with each pathway involving a series of core courses and elective courses. The first common pathway emphasizes

a progression of courses related to electrical systems and control technologies. The second common pathway reflects a sequence of courses that focus on circuit analysis and power system technologies.

Overall, Cluster 3 represents a group of students with a specific interest in power system-related subjects, particularly in the electrical engineering field. Their selective course choices and stronger performance in power system-related courses demonstrate a coherent academic profile. The findings from this cluster can provide valuable insights for curriculum planners and academic advisors to tailor course offerings and support services to meet the unique needs and interests of students with a preference for power systems engineering.

Among the six clusters, Cluster 4 stands out as an exceptional group of students within this engineering program, see Figure 8. With 268 students, it is the second-largest cluster, and it exhibits outstanding achievement in all subjects, consistently scoring above average across the board. The significant number of students in this cluster highlights the substantial presence of high-performing individuals within the student population.

Critical courses in Cluster 4, such as math8, Core4, and Core7, play a pivotal role in this cluster's outstanding performance. These courses are essential in providing students with a strong foundation in mathematical techniques, which are fundamental for understanding advanced engineering concepts. Core4 and Core7, in particular, are related to Signals and Circuits, and they require a solid grasp of mathematical principles from Math1 and Math2. The fact that students excel in these critical courses contributes significantly to their overall study performance, keeping it consistently above average.

A notable common pathway followed by students in Cluster 4 as shown in Figure 9 involves Math1, Core1a, Core4, and Core15 courses. Interestingly, this common pathway is mainly composed of first-year courses and are prerequisites for the later courses. This suggests that the success of students in Cluster 4 begins early in their academic journey, laying the groundwork for their achievements in subsequent years. Furthermore, an intriguing observation is that this group of students shows a preference for elective courses in High Voltage Engineering, which is more focused on electrical engineering rather than electronics. This choice indicates a specific interest or affinity for electrical-based subjects within the engineering program.

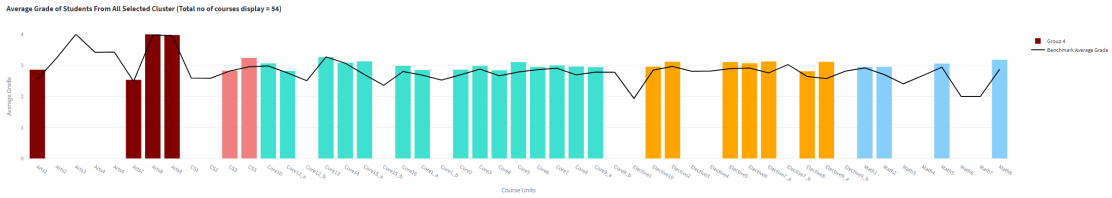


Figure 8: Bar chart depicting the average grades of students in Cluster 4 across various courses

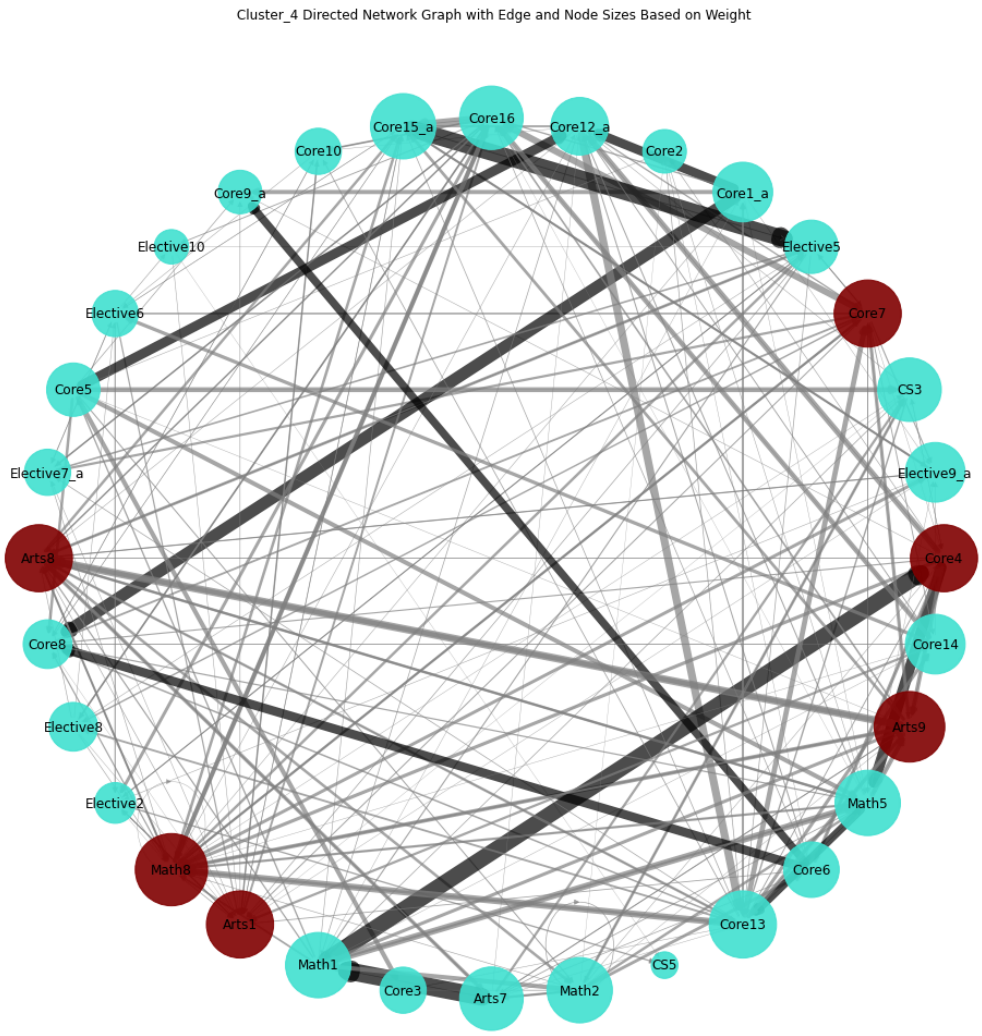


Figure 9: Network graph of Cluster 4

Overall, the remarkable performance of Cluster 4 students reflects their dedication, strong mathematical skills, and a solid understanding of core engineering concepts. The success of these students in critical courses, particularly those related to Signals and Circuits, highlights the importance of a strong foundation in mathematics and the value of excelling in core courses for academic excellence. Additionally, the early adoption of a common pathway with first-year courses underscores the significance of a solid start to a student's academic journey, setting the stage for future accomplishments in the engineering program.

Cluster 5 stands out as a distinct group in the analysis, comprising 45 students who face challenges in almost all the courses. Based on Figure 10, their performance is consistently poor, well below the average grade, across various categories of courses, including core courses, mathematics-related courses, and elective courses. This indicates that students in this cluster encounter difficulties in multiple subject areas, regardless of the course type. Upon further investigation, the clustering analysis reveals that most of the students in this cluster were enrolled in the university between 2008 and early 2013. This period could be significant in understanding the academic performance trends of this group and may be related to institutional changes or external factors that affected their learning experience.

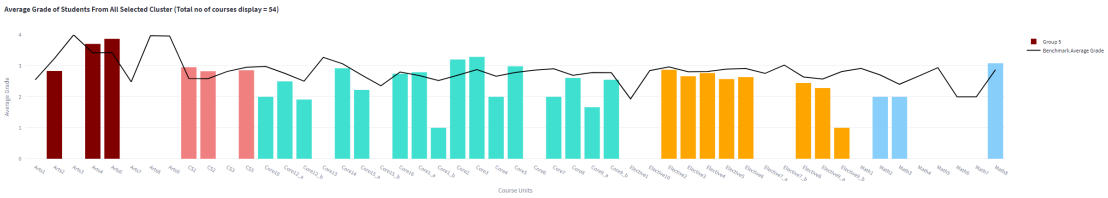


Figure 10: Bar chart depicting the average grades of students in Cluster 5 across various courses

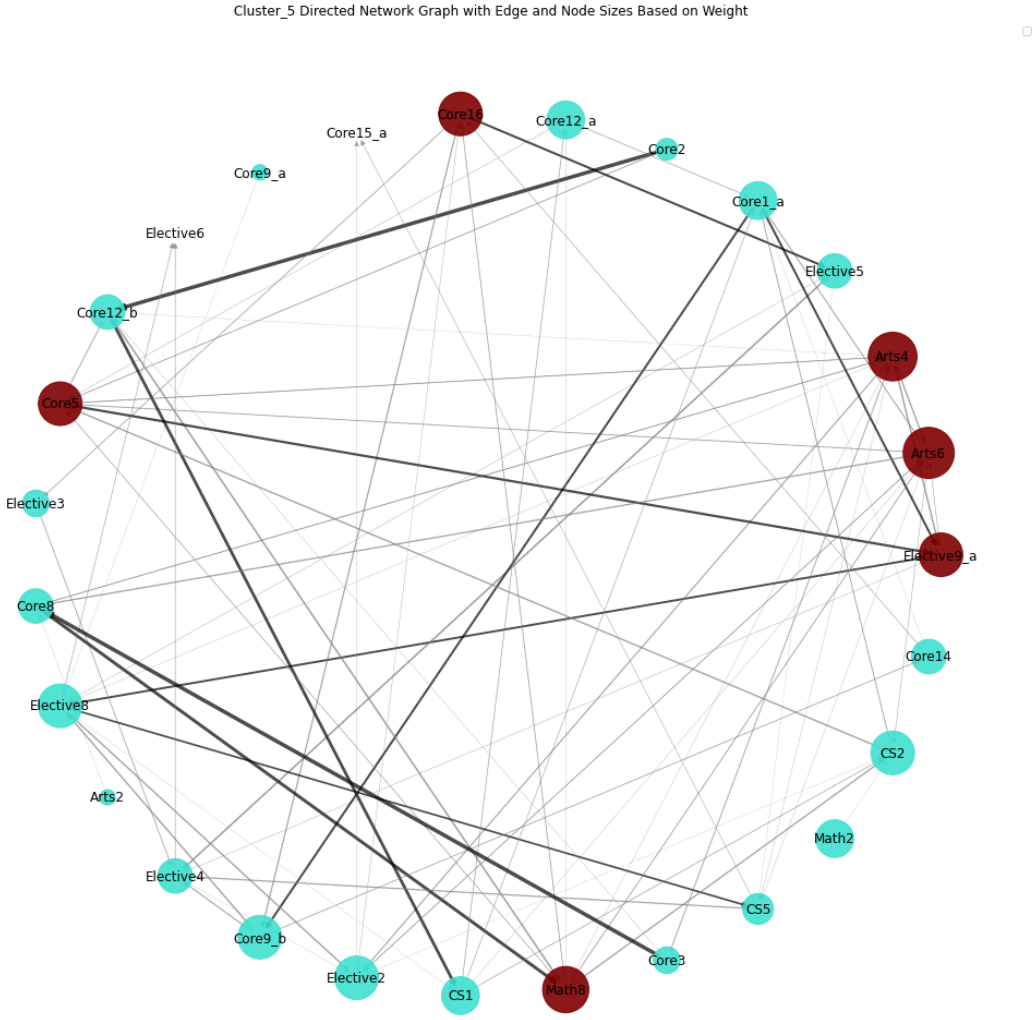


Figure 11: Network graph of Cluster 5

In terms of visualization, the nodes representing courses in the network graph of Figure 11 for Cluster 5 are relatively small, reflecting the smaller number of students in this cluster. The critical courses for students in this cluster are predominantly focused on math and digital electronics subjects, such as Core5, Core16, and Math8. This suggests that the challenges faced by these students may be particularly pronounced in these courses, which require a solid understanding of mathematical concepts and digital electronics principles. Interestingly, the elective courses chosen by most students in this cluster are Elective8 or Elective9. These elective choices align

with the critical courses that heavily involve digital electronics knowledge. Therefore, poor performance in these elective courses was expected, as these require a strong foundation in digital electronics, and struggling in foundation courses can have a ripple effect on the performance of these elective courses due to their interconnected concepts. The choice of these elective courses also indicated a preference for electronic-based subjects within this group.

Lastly, Cluster 0 and Cluster 2 share some similarities in terms of their characteristics, particularly in critical courses and academic performance, see Figure 12 and 13. Both Cluster 0 and Cluster 2 demonstrate a strong performance in liberty arts courses, with average GPAs close to 4.0. Liberty arts courses, which encompass subjects like language and moral studies, seem to resonate well with students in this cluster, leading to high achievements in these areas. This high level of achievement suggests that these students excel in humanities and social science-related subjects, where they demonstrate strong analytical and critical thinking skills. However, it is important to mention that their performance in the law for engineering course is below average. In addition to excelling in liberty arts courses, both cluster students achieve grades that are generally in line with the overall average across other various courses. This indicates that they are maintaining a consistent level of performance across different subject domains. Unlike some other clusters that may have strong weaknesses in specific subject areas, both of these clusters' students are neither significantly outstanding nor below average, making them a well-rounded group of students.

Besides, according to Figure 14 and 15, both clusters share common critical courses, including Math8, Core15_a, Arts4, and Arts6. These courses are considered fundamental to the academic curriculum and likely provide students with essential knowledge and skills that support their overall academic performance. The balanced performance suggests that students in these clusters possess a diverse set of academic skills and can adapt to various subjects effectively. These insights can provide valuable information to the institution to improve curriculum design and identify areas where additional support may be beneficial to help students excel further in their studies.

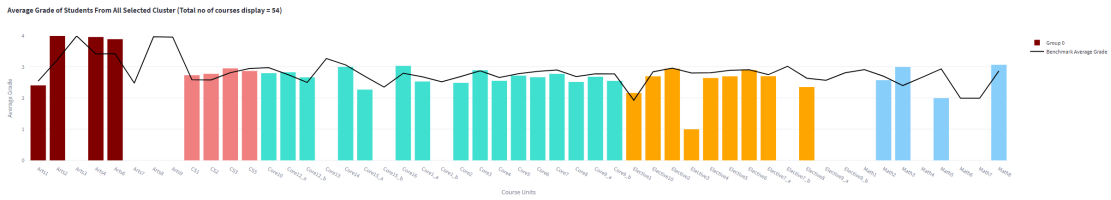


Figure 12: Bar chart depicting the average grades of students in Cluster 0 across various courses

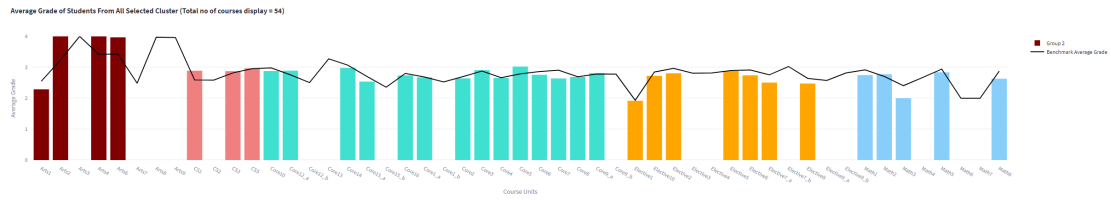


Figure 13: Bar chart depicting the average grades of students in Cluster 2 across various courses

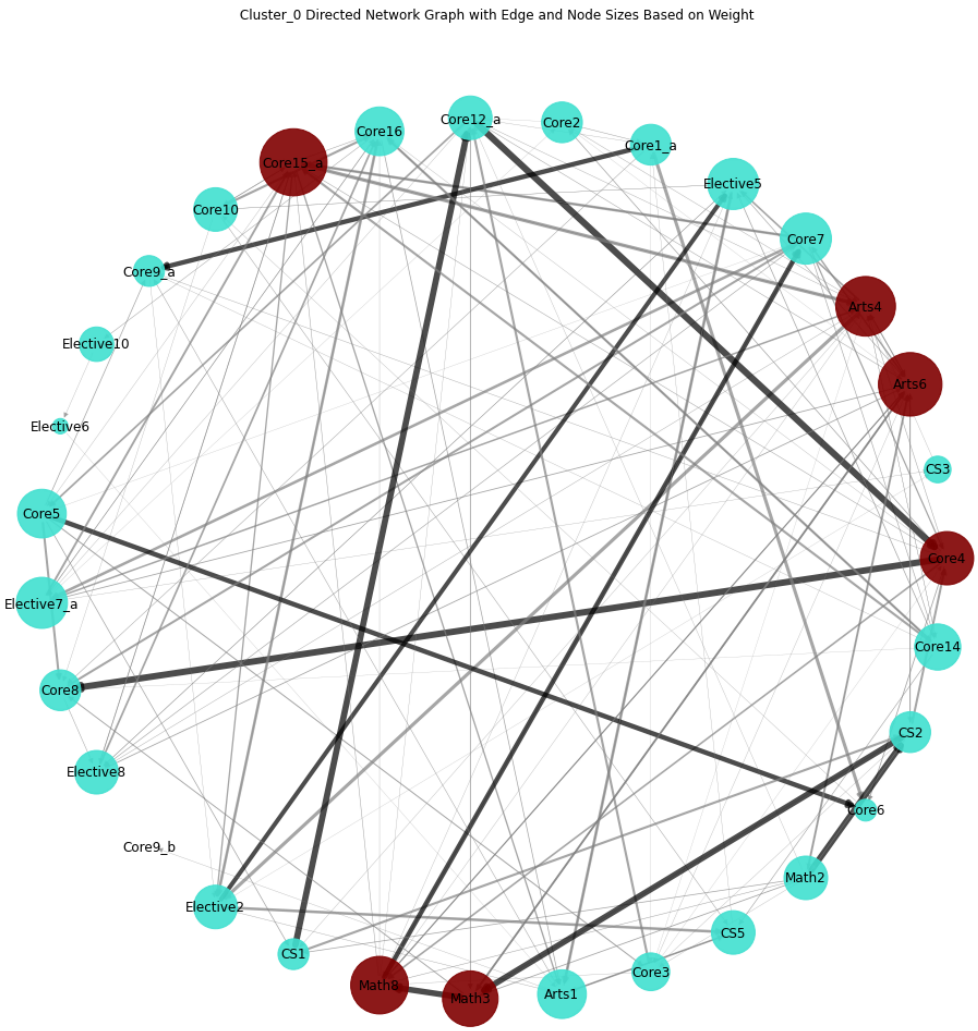


Figure 14: Network graph of Cluster 0

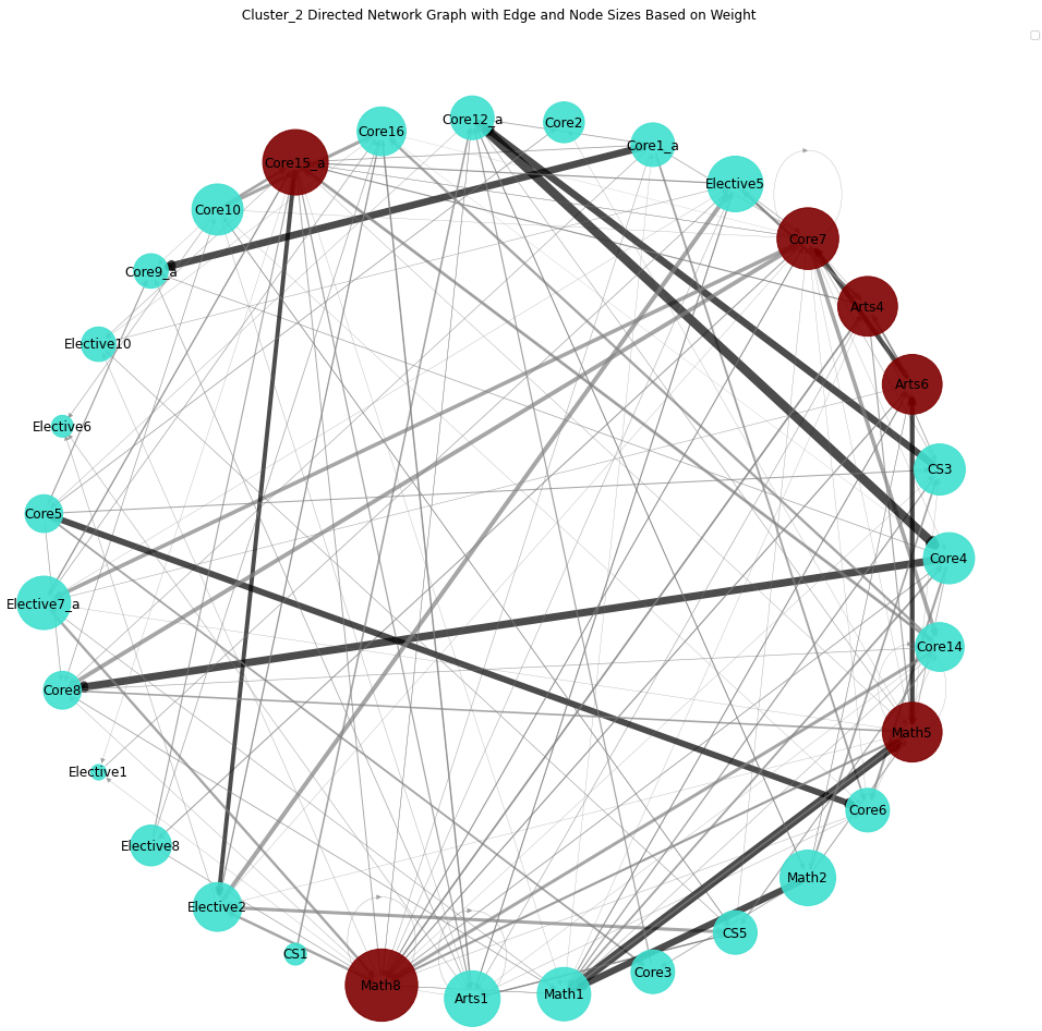


Figure 15: Network graph of Cluster 2

5. Conclusion

In conclusion, this study has demonstrated the value and potential of utilizing a unified learning analytic framework with educational data mining techniques in understanding students' academic performance and critical learning pathways. By employing a comprehensive framework that incorporates data collection, pre-processing, clustering analysis, and network analysis, the study has provided valuable insights into the factors influencing student success. The research findings highlight the importance of considering various elements, such as student performance, course components, and course sequences in analyzing student data. The application of unsupervised learning clustering algorithms has allowed for the identification of student clusters with similar academic performance and course patterns. Additionally, the construction of a course network and the analysis of network centrality have revealed critical nodes that significantly impact student outcomes and critical learning pathways that can inform curriculum design, instructional strategies, and intervention programs to enhance student success.

Looking ahead, future research can build upon this study by refining clustering algo-

rithms, incorporating temporal analysis, and expanding the framework to include additional data sources. Additionally, the development of intervention strategies and the assessment of their effectiveness could further enhance the practical applications of the framework.

Overall, this unified learning analytic framework serves as a powerful tool for educational institutions to comprehend the intricate relationship between academic performance and course pathways. It empowers educators to create data-driven strategies that address the unique needs of students, fostering a more inclusive and enriching learning environment. With this deeper understanding, institutions can strive to optimize their educational offerings and ensure that every student receives the necessary support to thrive academically and reach their full potential.

References

Abu Saa A. 2016. Educational data mining & students' performance prediction. *International Journal of Advanced Computer Science and Applications* **7**(5): 212–220.

Alhajj S., Alhajj A. & Özyer S.T. 2022. Combining multiple clustering and network analysis for discoveries in gene expression data. *ASONAM '21: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 502–509.

Almond-Dannenbring T., Easter M., Feng L., Guarcello M., Ham M., Machajewski S., Maness H., Miller A., Mooney S., Moore A. & Kendall E. 2022. *A Framework for Student Success Analytics*. EDUCAUSE Publications.

Bharara S., Sabitha S. & Bansal A. 2018. Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies* **23**(2): 957–984.

Borgatti S.P. & Everett M.G. 2006. A graph-theoretic perspective on centrality. *Social Networks* **28**(4): 466–484.

Caliński T. & Harabasz J. 1974. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods* **3**(1): 1–27.

Dol S.M. & Jawandhiya P.M. 2023. Classification technique and its combination with clustering and association rule mining in educational data mining — a survey. *Engineering Applications of Artificial Intelligence* **122**: 106071.

Govindasamy K. & Thambusamy V. 2018. Analysis of student academic performance using clustering techniques. *International Journal of Pure and Applied Mathematics* **119**(15): 309–322.

Greenacre M., Groenen P., Hastie T., Iodice D'Enza A., Markos A. & Tuzhilina E. 2022. Principal component analysis. *Nature Reviews Methods Primers* **2**(1): 100.

Grunspan D.Z., Wiggins B.L. & Goodreau S.M. 2014. Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE Life Sciences Education* **13**(2): 167–178.

Jain A.K., Murty M.N. & Flynn P.J. 1999. Data clustering: A review. *ACM computing surveys (CSUR)* **31**(3): 264–323.

Jan S.K., Vlachopoulos P. & Parsell M. 2019. Social network analysis and learning communities in higher education online learning : A systematic literature review. *Online Learning* **23**(1).

Joshi A., Desai P. & Tewari P. 2020. Learning analytics framework for measuring students' performance and teachers' involvement through problem based learning in engineering education. *Procedia Computer Science* **172**: 954–959.

Khalil M., Prinsloo P. & Slade S. 2022. A comparison of learning analytics frameworks: a systematic review. *LAK22: 12th International Learning Analytics and Knowledge Conference*, pp. 152–163.

Križanić S. 2020. Educational data mining using cluster analysis and decision tree technique: A case study. *International Journal of Engineering Business Management* **12**: 1847979020908675.

Leuschner J., Schmidt M., Fernsel P., Lachmund D., Boskamp T. & Maass P. 2019. Supervised non-negative matrix factorization methods for MALDI imaging applications. *Bioinformatics* **35**(11): 1940 – 1947.

Omar T., Alzahrani A. & Zohdy M. 2020. Clustering approach for analyzing the student's efficiency and performance based on data. *Journal of Data Analysis and Information Processing* **8**(3): 171 – 182.

Samanta P., Sarkar D., Kole D.K. & Jana P. 2021. Social network analysis in education: A study. *Computational Intelligence in Digital Pedagogy* ((eds.). A. Deyasi, S. Mukherjee, A. Mukherjee, A. K. Bhattacharjee & A. Mondal), pp. 65 – 83. Singapore: Springer.

Saqr M., Fors U., Tedre M. & Nouri J. 2018. How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PLoS One* **13**(3): e0194777.

Shahapure K.R. & Nicholas C. 2020. Cluster quality analysis using silhouette score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 747–748.

Su Y.S., Lin Y.D. & Liu T.Q. 2023. Applying machine learning technologies to explore students' learning features and performance prediction. *Front Neurosci* **16**(3): 1018005.

Lee Kong Chian Faculty of Engineering & Science
Universiti Tunku Abdul Rahman
43200 UTAR Kajang
Selangor, MALAYSIA
E-mail: tanyenlyn97@gmail.com, gohyk@utar.edu.my, laiac@utar.edu.my, ngeowym@utar.edu.my*

Received: 5 September 2023
Accepted: 18 January 2024

*Corresponding author