NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

OPTIMIZING ADAPTIVE LEARNING USING STATISTICAL AND NETWORK ANALYSIS WITHIN THE CHUNK LEARNING SYSTEM

by

Welvinjohn Lucero

June 2020

Thesis Advisor: Ralucca Gera
Co-Advisor: Michelle L. Isenhour

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2020	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE OPTIMIZING ADAPTIVE LEARNING USING STATISTICAL AND NETWORK ANALYSIS WITHIN THE CHUNK LEARNING SYSTEM			**5. FUNDING NUMBERS**
6. AUTHOR(S) Welvinjohn Lucero			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

What can we learn from analyzing how a learner operates within an e-learning system? This research examines an adaptive education system known as CHUNK Learning. This system converts educational material into a network composed of nodes, or CHUNKs of educational content, and edges that represent the connections between some of the nodes. As of now, learners have freedom of maneuver to navigate within the system at their leisure. Each of their actions is a piece of data that can be used by an instructor to comprehend whether a learner is effectively learning or not. The CHUNK Learning system has not yet utilized this valuable data to improve the complex teaching-learning process that occurs in an e-learning environment. We propose a solution to this problem by utilizing user analytics based on two criteria: number of completed educational modules and the number of content views. We conduct two different mathematical approaches based on statistical analysis and network science that allow for a thorough analysis of user data to determine vital trends that enhance the situational awareness of CHUNK Learning. We look to determine user competency scores that may reveal troubling areas of deficiency which may enable instructors to tailor their teaching methods to address each user's specific needs. In addition, we can further personalize learning to meet user needs by determining the optimal learning content for each course.

14. SUBJECT TERMS electronic learning, personalized learning, adaptive learning, statistical theory, network theory, cognitive theory			**15. NUMBER OF PAGES** 81
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

OPTIMIZING ADAPTIVE LEARNING USING STATISTICAL AND NETWORK ANALYSIS WITHIN THE CHUNK LEARNING SYSTEM

Welvinjohn Lucero
Captain, United States Army
BS, U.S. Military Academy, 2010

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED MATHEMATICS

from the

NAVAL POSTGRADUATE SCHOOL
June 2020

Approved by: Ralucca Gera
Advisor

Michelle L. Isenhour
Co-Advisor

Wei Kang
Chair, Department of Applied Mathematics

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

What can we learn from analyzing how a learner operates within an e-learning system? This research examines an adaptive education system known as CHUNK Learning. This system converts educational material into a network composed of nodes, or CHUNKs of educational content, and edges that represent the connections between some of the nodes. As of now, learners have freedom of maneuver to navigate within the system at their leisure. Each of their actions is a piece of data that can be used by an instructor to comprehend whether a learner is effectively learning or not. The CHUNK Learning system has not yet utilized this valuable data to improve the complex teaching–learning process that occurs in an e-learning environment. We propose a solution to this problem by utilizing user analytics based on two criteria: number of completed educational modules and the number of content views. We conduct two different mathematical approaches based on statistical analysis and network science that allow for a thorough analysis of user data to determine vital trends that enhance the situational awareness of CHUNK Learning. We look to determine user competency scores that may reveal troubling areas of deficiency that may enable instructors to tailor their teaching methods to address each user's specific needs. In addition, we can further personalize learning to meet user needs by determining the optimal learning content for each course.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

AES	adaptive educational system
CBE	competency-based education
CHUNK	Curated Heuristic Using a Network of Knowledge
CR	completed CHUNKs/required CHUNKs
DoD	Department of Defense
e-learning	electronic-learning
GUI	graphical user interface
NPS	Naval Postgraduate School
SD	standard deviation
STAR	standardized algorithm of required CHUNKs
VC	views of completed CHUNKs/completed CHUNKs
STEM	science, technology, engineering, and mathematics

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Today's students are more exposed to technology in the classroom than their predecessors. The use of technology has become a way of life for modern society yet traditional pedagogy in a brick-and-mortar institution still remains the norm. Recent studies point toward an improved learning experience for students when e-learning is involved. E-learning can offer instructors a systematic and adaptive process to learn valuable details about their students and gain the versatility to model their courses using student input.

This thesis examines data from the e-learning system named CHUNK Learning. This system builds a network of individual CHUNKs that each contain a plethora of learning content such as videos, sample code, research papers and PowerPoint presentations. Currently, CHUNK Learning is a platform that uses learners' profiles through the information provided by the user. While this information is useful, users may not accurately fill out their profile or even update them after initially creating them. Therefore, we propose a dynamic approach using a mathematically-based algorithm to provide an analysis of the content users studied in order to guide the faculty that utilize CHUNK Learning in support of their courses. Additionally, through modeling user interactions as networks, we also introduce a second algorithm that allows us to discover which educational modules in a course are most representative of that course.

The first algorithm is based on a statistical analysis of CHUNK user data that measures two particular user attributes: CHUNK completion rates and completed CHUNK views. This algorithm utilizes the z-score standardization method which enables us to place each user into a specific group based on their z-scores related to these two attributes. The output of the algorithm assigns each user a competence level that can be used by instructors to make personalized recommendations to their students that may improve the teaching-learning process.

The second algorithm utilizes a network science approach that analyzes the connections between users and completed CHUNKs to build an individual network within each course. We use tools of network science to determine the most important selec-

tion of CHUNKs based upon the results of multiple parameters.

The combined user-centric approach within CHUNK Learning may serve as a model for other e-learning systems to follow. While the course data was limited to graduate students from NPS, we may extract additional critical feedback from applying these algorithms to students at all education levels.

Acknowledgments

First, I am extremely thankful for having two dedicated advisors, Professor Ralucca Gera and Professor Michelle Isenhour, who offered their expert guidance and invaluable mentorship at each step of the way. Ralucca ensured that I possessed the necessary knowledge to discuss both Graph Theory and Network Science, while Michelle spent countless hours of her personal time verifying and assisting with the statistical analysis portion of my thesis. I am also very grateful for her instruction regarding the different software programs to include Gephi, CoCalc/Python, and Microsoft Excel.

Next, I would also like to thank the many professors and students in the Applied Mathematics and Operations Research departments who have helped me become a better mathematician. Special thanks to my West Point cohort peers for all the good times and laughs while persevering through some challenging quarters at NPS together.

Last but not least I would like to thank my wife, Karina, for her unwavering dedication and support during these last two years. I am absolutely blessed to have you by my side and could not ask for a better partner in crime, especially during these trying times. Much of the work that was put into this thesis would not be possible without your understanding and patience.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

1.1 Motivation

In a world dominated by globalization and advancing technologies, learners are exposed to various pedagogies, each with their own set of advantages and disadvantages. Instructors possess an innate responsibility to ensure the best possible learning environment for their students and must strive to improve their methods each day. Most educational institutions today implement some form of traditional education where student success remains highly dependent upon how well each student adapts to their instructor's teaching style. By exploring the use of non-traditional pedagogies, instructors can incorporate new technologies to optimize the teaching-learning process.

1.2 Problem Statement

Researchers at the Naval Postgraduate School (NPS) in Monterey, CA, USA, developed a competency-based electronic-learning (e-learning) system, known as Curated Heuristic Using a Network of Knowledge (CHUNK) for Continuum of Learning, that allows students to conduct learning at their own pace in a non-linear manner [1]. This is done through the CHUNK Learning environment which emphasizes a personalized and adaptive approach to education. The latest system relies on students to create and update a user profile that allows learning to be modeled upon personal attributes such as educational background, skills and interests. The system works well until current students fail to regularly update their profile. Although CHUNK Learning aims to provide a tailored learning path for each student in this manner, there are other methods that can be used to match content that is specific to each student. Very often, traditional pedagogy dictates that students must adapt to their instructor's teaching philosophy in order to adeptly understand a given subject. Though this methodology may be useful for some students, there are others who might suffer either due to lack of clarity from the instructor or the arduous nature of the material at hand. These students may be ill-equipped to join a 21st century workforce that "must be able to address dynamic situations, think critically, and solve problems by accessing and analyzing information" [2]. If instructors are able to better understand their students especially how they approach learning, then surely the teaching-learning process would lead to a more pleasant and rewarding experience for both parties. Students would likely buy-in to this type of approach when instructors effectively respond to their feedback. Therefore, we are motivated to incorporate a bottom-up approach to e-learning, where student analytics point out key attributes of a student to provide instructors with clear feedback on areas of concern surrounding the educational content. We aim to identify students by their competence levels in a course at any given time.

Moreover, instructors are often faced with situations where they have limited time to cover an entire course or they meet students who are exploratory learners that may express interest towards a couple of topics within the course. This is an interesting problem that often leads instructors to arbitrarily choose topics that they believe are the best representation of the course material based on their experience. The CHUNK Learning system possesses the versatility to recommend particular CHUNKs given its interconnected structure. We can use network science analysis to provide clarity on which topics may be the most important based on student interactions in the CHUNK Learning system.

1.3 Purpose

In this project, we develop a statistical analysis algorithm that is dependent upon student data in CHUNK Learning to generate a score for each student based on competency and interest. This algorithm provides the potential for instructors to have a snapshot of each student's learning capability. We specifically look at two areas of student analytics: content completion rate and content views. This data provides us with critical feedback that allows instructors to gauge where each of their students are at any given point in the material. In doing so, instructors can facilitate a student's personalized path to identify whether he or she needs to spend more time on a particular topic or would benefit from learning supplementary content to gain additional expertise.

The second objective is to conduct a network science analysis of the connections between students and the educational content. We believe that the analysis will reveal a course's highest priority topics based on network science parameters. This unique perspective will improve the CHUNK Learning system's ability to meet the competing demands of students who are interested in learning about subjects outside of their required coursework but within a limited time frame. Ultimately, the analysis may provide future students a personalized and efficient package of content to learn.

1.4 Thesis Structure

We have organized this thesis into five chapters to include Introduction, Background, Data and Methodology, Results and Analysis, and Future Work and Recommendations. In Chapter II, we describe important background information highlighting the various theories that dominate our project. Next, in Chapter III, we outline the data and the methodologies necessary to develop each of the two algorithms. In Chapter IV, we present the results and findings of our statistical analysis and network science analysis. Finally, we give our recommendations for future work and conclusions in Chapter V.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Background

Within this chapter, we investigate a variety of concepts that must be explained before continuing further. We provide key findings from various fields of study applicable to e-learning and introduce important definitions that formulate the base of our mathematical models. Specifically, we will discuss key terminology relating to the following topics: adaptive learning, competency learning, statistics, graph theory, network science, and CHUNK Learning.

2.1 Adaptive Learning Theory

A traditional learning environment has been the established norm for centuries with key advantages such as valuable face-to-face mentorship between students and instructors, and group projects where students work together and learn from each other. Yet, with society's growing dependence on technology, there is an abundance of research indicating that e-learning solves many of the issues that impact a traditional learning atmosphere [3]. E-learning harnesses today's technologies to develop a viable educational package that can used by anyone who is connected to the internet. E-learning is the act of utilizing electronic devices like cell phones and tablets to "support individual learning or organizational performance goals" [3]. This learning method has gained favor in developing countries such as India and Malaysia due to its key advantage of affordability [4]. In addition, e-learning greatly enhances a classroom environment through the introduction of supplementary resources. Instructors can take full advantage of e-learning to implement a dynamic educational approach such as blended learning, an approach that combines traditional learning and the internet to create an enhanced learning experience for students. Further, organizations rely on e-learning to reduce travel expenses and training time involved with traditional face-to-face learning [3]. Research has demonstrated that "information and communication technology-based interventions generally resulted in positive impact over traditional learning" but these gains have failed to reshape the world's perspective on standard pedagogies [4]. Thus, we introduce adaptive learning, an important concept that utilizes e-learning and may fundamentally change how we view modern-day teaching and learning.

Adaptive learning is the dynamic process where e-learning shapes content based upon "student performance" [5]. The advancement of technology has led to the development of complex systems, capable of gathering refined data which utilize different learning analytics to identify the correct approach to a student's learning path. Adaptive learning is able to provide effective learning by supporting diverse learning paths and materials to fit learners' unique needs and lifestyles [6].

An adaptive educational system (AES) combines e-learning and adaptive learning together with the fundamental principle that all learners are unique. Therefore, we must account for their different attributes to allow for seamless communication and instruction between instructor and student. These attributes may include each learner's depth of knowledge on a particular subject, their preferred style of learning, and other cognitive skills they might possess [4]. Thus, AESs must account for a large number of variables relating to the characteristics of each and every user. When applied to a class of diverse students, this process becomes increasingly complex.

Among various individual characteristics, learning styles are considered to be important factors in e-learning, and due consideration of these factors is necessary in delivering a quality learning experience to learners. Manochehr found that if learning content was successfully linked to student's learning styles, then the quality of learning may be improved [7]. Yet, several new learning style-based AESs were developed to deliver adaptive learning to students and their results have been contentious, unable to positively impact academic achievement [4]. While learning styles are critical to understanding different types of learners, we must not solely consider one single aspect of student learning when creating an AES, as this alone cannot make a profound difference towards enhancing the teaching-learning process. Therefore, the development of AESs must aim to combine learning styles with other parameters to provide a much fuller and adequate adaptive learning experience to the learner. Thus, we introduce one of those cognitive parameters, working memory.

Baddeley defines working memory as "a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning" [8]. It is not to be confused with short-term memory which refers only to momentary storage of any information being processed. Additionally, individuals who possess a high working memory capacity are not necessarily more intelligent than those individuals with a low working memory capacity. Yet, there have been encouraging studies which support the claim that working memory is directly correlated to a person's fluid intelligence [9]. In the context of an AES, instructors will need to compensate for their students' varying levels of working memory capacity, where certain academic topics may overload some students, discouraging them from fully engaging in the learning process [4]. Tsianos et al. suggest that students with low working memory capacity can improve to levels of medium or high working memory capacity when their instructors are able to tailor content to match their capacity [10]. Ultimately, this research suggests that accounting for cognitive measures such as learning styles and working memory can potentially enrich the learner's experience in an AES such as CHUNK Learning.

2.2 Competency Learning Theory

Moving into the 21st Century, the rigid demands of workplaces have transitioned from one of physical inputs towards a knowledge-based economy that emphasizes the intellect of employees [11]. This certainly indicates that companies will seek to hire viable employees who possess specialized training and multiple certifications. It seems reasonable that these employees would benefit greatly from competency-based education (CBE), a pedagogy that ensures students can properly demonstrate what they have learned and understand the different applications it may have [11]. CBE, as described by Spady, is an adaptive process that properly instructs, measures and certifies an individual in a flexible time frame to determine if he or she has met competency-driven learning outcomes [12]. Instructors use CBE to introduce a more comfortable learning experience to their students, one in which they can learn at their own pace. Self-directed learning provides students with the freedom to explore and choose their own education needs. CBE prioritizes a student's time, allowing them to learn exactly when they want to. CBE programs aim to improve various skills such as studying habits, use of implicit knowledge and critical thinking that effectively prepare students for real-world challenges [2]. According to Gardner, the goal of CBE programs is purely to indicate a learner's competency by "written or other demonstration of knowledge of a subject or a particular skill, often those reflecting workplace experiences and responsibilities" [11]. This enables students to navigate through content at their own pace based on their demonstration of the required knowledge. Students enrolled in CHUNK Learning would surely benefit from the system knowing exactly what their competence level is in a given topic. This nests with the CHUNK Learning system's goal to assess users through the use of competency-based assessments.

2.3 Statistical Theory

The study of statistics provides the foundation for how we must observe and interact with the data we collect. This discipline enables us to make smart judgments and enhance our decision-making when presented with situations of uncertainty and variation [13]. Statistics equip us with essential tools to collect vital information and develop findings. Essentially, we can make inferences from our data to generalize across differing populations. Among the various visual techniques that can be used in statistics, we are interested specifically in displaying *histograms* for our discrete data. A histogram is a type of plot that presents the frequency distribution of a set of data that is usually either discrete or continuous. The data in this project is discrete so we construct a histogram by "subdividing the measurement axis into a suitable number of classes, such that each observation is contained in exactly one class" [13]. We also refer to these classes as *bins*, and we will look to distribute our data into equal bin widths.

We now list a number of terms that will be referenced throughout the thesis. We start with this definition of *variance*.

Definition 2.3.1 *Variance*

> *Variance, denoted by* σ^2*, is given by* $\sigma^2 = \frac{\sum(x_i-\mu)^2}{n}$*, where* $x_i - \mu$ *is the deviation of the* i^{th} *observation from the population mean,* μ*, and* n *is the size of the population [13].*

As can be seen from Definition 2.3.1, the variance is simply the average of the squared deviations from the population mean. It is used to measure the variability in a given population. From the variance, we are able to derive the standard deviation.

Definition 2.3.2 *Standard Deviation*

> *Standard deviation, denoted by σ, is the square root of the variance, $\sigma = \sqrt{\sigma^2}$ [13].*

The standard deviation can be interpreted as the size of a typical or representative deviation from the population mean within a given set of data. Within CHUNK Learning, computing the standard deviation allows us to obtain a measure representing how much a typical learner's usage of CHUNK learning differs from the average user within each course. It is the unit of length upon which this project's statistical analysis algorithm depends on.

Within data sets, there are multiple variables that have different measurement standards. Statistical tools, like the computation of a z-score, enable us to establish a standard among these variables in order to compare performance and conduct analysis among slightly differing populations.

Definition 2.3.3 *Z-score*

> *Z-score is the number of standard deviations by which the raw score's value is above or below the mean value [13]. The equation is given by:*

$$Z = \frac{x - \mu}{\sigma}$$

We can utilize the z-score as a means to effectively standardize the data we will use in this project since we must be able to interpret the data for differing courses, each with a contrasting number of students and learning requirements.

2.4 Graph Theory and Network Science

In this section, we discuss new mathematical concepts, graph theory and network science, that allow us to represent data as interconnected graphs and networks.

2.4.1 Graph Theory

Graph theory is one of the newest fields within mathematics, first arising in the early 18th century, and did not become an official branch of mathematics until the late 19th century [14]. Despite its infancy, this area of study allows for a different perspective when studying a multitude of different problems. Graph theory introduces the use of graphs as a means to model the relationships that exist in nature, allowing for connections to develop in places that may surprise us. When applied to an AES, we are left with an entirely unique and more holistic perspective than most mathematical applications, to include statistics, offer. Graph theory allows us to understand the essential framework needed to employ the vast array of tools contained within network science that are robust enough to conduct a comprehensive quantitative analysis. Thus, we list the following terms and their associated definitions to provide a basic understanding of graph theory and how it is applied to this thesis.

Definition 2.4.1 *Graph*

> *A graph G consists of a finite nonempty set V of objects called vertices (the singular is vertex) and a set E of 2-element subsets of V called edges. The sets V and E are the vertex set and edge set of G, respectively. So a graph G is a pair (actually an ordered pair) of two sets V and E. For this reason, some write $G = (V, E)$ [14].*

There is much interest that may arise when building graphs from the data collected in a learning system, especially an e-learning system where data is easily accessible and usually quite robust. In these e-learning graphs, the *vertices*, also known as *nodes*, represent the set of students and the learning content they each participate in, whereas the *edges* are the 2-element subsets that join the students and the learning content together. As we look particularly closer at nodes and edges, we may be drawn to their particular *degree*.

Definition 2.4.2 *Degree*

> *The degree of a vertex v in a graph G is the number of edges incident with v and is denoted by $\deg_G v$ or simply by $\deg v$ if the graph G is clear from the context. Also, $\deg v$ is the number of vertices adjacent to v [14].*

The graph theory term, *degree*, is an important parameter that provides valuable information about the data set used in the project. When applied to CHUNK Learning, high degree users are those who have completed a high number of CHUNKs. On the contrary, users are deemed as having low degree if they completed a low amount of CHUNKs. This parameter will be largely influential during the network analysis since nodes with high degree may be more correlated to importance when compared with nodes of lesser degree.

2.4.2 Network Science

The ideology we discussed in graph theory now gives us the opportunity to introduce network science, the mathematical approach we will take to transform graphs into networks. Newman first described network science as a way to interpret the complex "behaviour of real-world networked systems" [15]. Definition 2.4.1 can be modified to interpret complex networks, rather than purely laying focus on the relations between nodes and edges. Simply, networks are much more versatile than graphs, as they allow us to model different phenomena that exist in the world from biological to social phenomena. Newman offered this definition of a network:

Definition 2.4.3 *Network*

> *A network is a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else [15].*

While the primary view of educational content within the CHUNK Learning system itself is a large network, we are able to make smaller networks out of the individual courses that exist within this larger network. Moreover, network science possesses data analysis tools to analyze and extract valuable connections that exist within the network among the various nodes and edges. As network science looks to analyze the data through a network, one of the most common measures is the identification of influential nodes within the scope of the network. We use the term, centrality, to denote the most important nodes contained in the entire network. Two of the most commonly used types of centrality measures include *eigenvector centrality* and *betweenness centrality*, which will be used for this project.

Definition 2.4.4 *Centrality*

> *Eigenvector centrality is the principal eigenvector of the adjacency matrix defining the network. Betweenness centrality is the share of times that a vertex i needs a vertex k (whose centrality is being measured) in order to reach a vertex j via the shortest path [16].*

For this project, we are interested in identifying the CHUNKs that have high centrality because influential educational content may significantly impact a user's experience with the CHUNK Learning system. As a result, these influential nodes can inform the CHUNK Learning system as to which CHUNKs to prioritize, enhancing the organization and layout of the material. Next, we explore the usefulness of the *k-core*, a subgraph that possesses useful properties when fully analyzed. Unlike centrality which looks at the importance of single nodes, the k-core examines subgraphs or sets of nodes that together are more important than the rest of the subgraphs within the network.

Definition 2.4.5 *k-core*

> *A maximal connected subgraph, where the elements of the subgraph are connected to at least k other elements of the same subgraph; alternatively: the union of all k-shells with indices greater or equal to k, where the k-shell is defined as the set of consecutively removed nodes and belonging links having a degree $\leq k$ [17].*

For our research, we will be using the concept of the k-core to determine valuable subgraphs that exist within the CHUNK Learning network that allow us to pinpoint the group of the most influential educational content presented to users. We aim to obtain a core-periphery structure for each subgraph to uncover core nodes and differentiate them from less important periphery nodes. According to Csermely, "the network core can be regarded as a highly degenerate segment of the complex system, where the densely intertwined pathways can substitute and/or support each other" [17]. Through the use of k-core analysis, we can identify the core CHUNKs of a specified course in the CHUNK Learning network to aid both instructor and learner in the teaching-learning process.

2.5 CHUNK Learning

We now introduce the CHUNK Learning project in more details, first its interface and then the learning process.

2.5.1 The CHUNK Learning Interface

This section reveals important terminology specific to the system we use during this project. We use CHUNK Learning terms to explain the breakdown of educational material into individual nodes that facilitate an adaptive learning approach. The CHUNK Learning Explorer (Figure 2.1) is the system's graphical user interface (GUI).

In this figure, the key provided on the left hand side of the interface displays many of the important terms that make up the anatomy of the CHUNK Learning system. The largest nodes in the system are called topics. Each topic represents a single educational course encompassing one academic term, usually either a quarter or a semester. Therefore, we use the terms topic and course interchangeably. Next, each topic is divided further into individual units, similar to chapters within a course's textbook. These units contain one or more CHUNKs, the building blocks of the CHUNK Learning interface. Moreover, the CHUNK is broken further down into four separate CHUNKlets that model a "Why-How-Methodology-Assessment format" [1]. Finally, each CHUNKlet contains one or more Activities. Cleven describes these activities as the smallest form of educational content that learners interact with which may

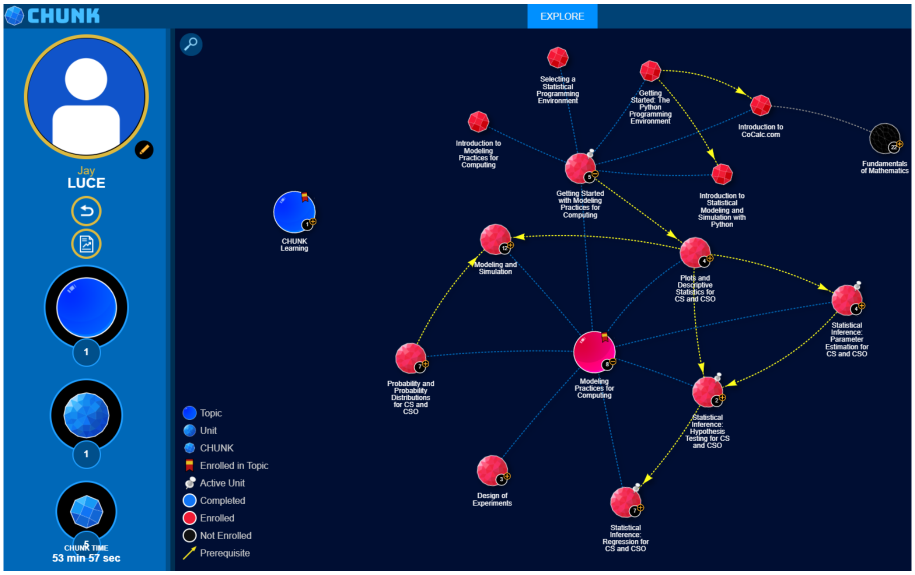


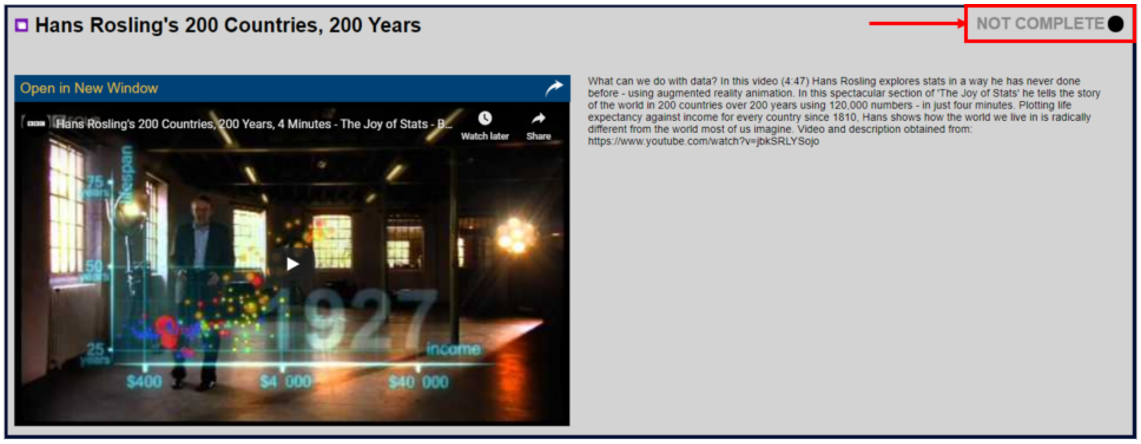
Figure 2.1: CHUNK Learning Explorer interface

content items such as "videos, webinars, codes, games, articles, [or] assessments" [18].

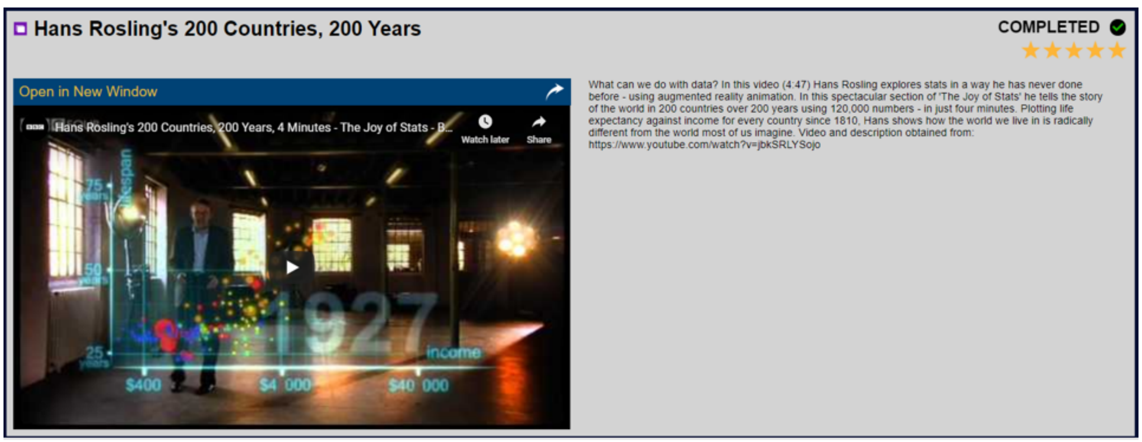
2.5.2 The CHUNK Learning Process

The goal of CHUNK Learning is to provide a customized experience for each learner where their needs and skills are fully taken into consideration. This system diverges from the traditional education model, optimizing the content in a manner that leverages a learner's learning styles and problem-solving methods. The experience between a student and an instructor is significantly enhanced through various modes of feedback which act to effectively process and optimize the educational material [1]. Learning is facilitated through the completion of various CHUNKlets within a single CHUNK. Since the CHUNK Learning system is a functioning prototype and not a full-scale operating system, some features are not optimal. One of those features is the manner in which students complete content. As a learner completes the activities in a CHUNKlet, he or she must manually click on the completion checkbox as seen in Figure 2.2.

A CHUNKlet may contain a single activity or multiple activities. Upon completing the last activity within a CHUNKlet, the system credits the learner with a completed CHUNKlet. As previously discussed, a CHUNK contains multiple CHUN-



(a) An incomplete Activity



(b) A completed Activity

Figure 2.2: Process of completing an Activity within a CHUNKlet in CHUNK Learning

Klets arranged in a "Why-How-Methodology-Assessment" format. Figure 2.3 shows the various "Why-How-Methodology-Assessment" components of a CHUNK as displayed using the CHUNK Learning interface.

Generally, the "Why" CHUNKlet consists of one or more video or reading activities that explain the general importance behind a specific topic. The intent of this CHUNKlet is to arouse interest from the student by providing a short and insightful motivation to learn. The "How" CHUNKlet is comprised of one or more 3-5 minute video or reading activities that detail the applications of a topic. Students may find that this CHUNKlet provides a more in-depth understanding of the topic in an en-

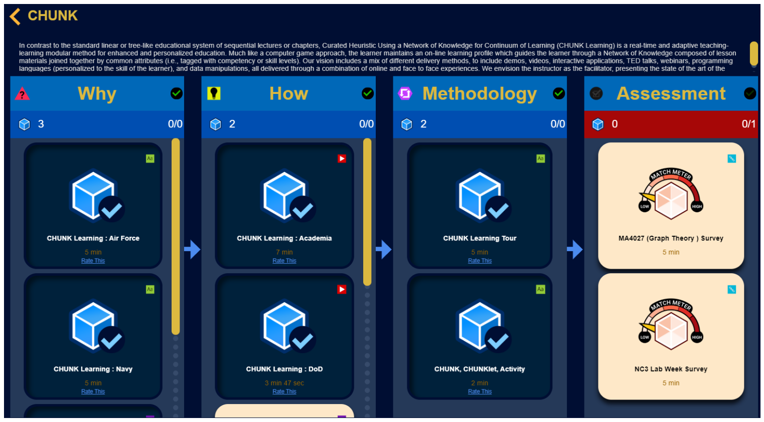


Figure 2.3: Various components of a CHUNK

gaging and entertaining manner. Next, the "Methodology" CHUNKlet is the heart of the CHUNK and contains all the relevant instructional material ranging from reading assignments and video lectures to PowerPoint presentations, coding exercises, and practical exercises [19]. This CHUNKlet is where the learning happens, and the instructor can truly tailor content in a feasible manner that is best suited for their students. Lastly, the "Assessment" CHUNKlet is used to evaluate student learning or skill acquisition at the conclusion of a CHUNK. This CHUNKlet typically consists of a summative assessment activity including tests, quizzes, homework assignments, laboratory exercises, or research projects [19]. Any assessment that instructors would use in a traditional classroom setting can be adapted to an "Assessment" CHUNKlet.

Most CHUNKlets are optional, providing the learner the flexibility to explore as much or as little of the material based on his/her interests. Therefore, the successful completion of any one "Assessment" CHUNKlet will reward the student with a completed CHUNK. However, instructors are also able to make "Why", "How", or "Methodology" CHUNKlets mandatory. In these cases, the user must complete all activities within all mandatory CHUNKlets as well as one "Assessment" CHUNKlet before the system will reward the student with credit for a completed CHUNK. Within the CHUNK Explorer, the color of the CHUNK will change colors from red to blue to denote that the CHUNK is complete. For the purposes of this study, we identify three different types of CHUNKs in the system: required CHUNKs, recommended CHUNKs, and optional CHUNKs. A required CHUNK consists of learning content that is necessary to achieving the instructor's learning objectives for a specified course. The instructor identifies these CHUNKs on the course syllabus. A recommended CHUNK, also identified by the instructor as relevant to the course, contains learning content that acts as supplementary material the user may choose to complete for expanded learning. An optional CHUNK is any CHUNK that is available within CHUNK Learning but completely outside the scope of the current course that the user is enrolled in.

The exploration of the various topics in this thesis provide the necessary framework to create algorithms that allow us to interpret the meanings behind various user actions in the CHUNK Learning environment. The research on adaptive learning and competency learning establish the purpose for conducting an analysis of CHUNK Learning as these educational concepts become increasingly known and utilized throughout the world. We then incorporate the mathematical tools within statistics, graph theory and network science to develop each of these algorithms and provide the necessary analysis to reveal key trends among users enrolled in CHUNK Learning. The following chapter explains the CHUNK Learning data we use and the methodology for the statistical analysis algorithm and the network science analysis algorithm.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Data and Methodology

This chapter describes the data obtained from the CHUNK Learning network. Additionally, we discuss the methodology and the algorithm involved in the statistical analysis approach. Then, we reveal the methodology of the network science analysis approach using the network analysis and visualization program, Gephi [20].

3.1 CHUNK Learning Data for our Analysis

We collected data and performed our analysis on a total of 68 CHUNK Learning users enrolled in three different courses at NPS.

We perform our analysis on the following three data sets: MA4027 - Graph Theory, OS3307 - Modeling Practices for Computing, and OS3604 - Statistics and Data Analysis. This data was sourced directly from the CHUNK Learning system and gathered for the entire academic quarter in 2019 and 2020. All three courses utilized a flipped classroom learning approach for portions of the course, in which students are tasked with completing the course's assigned CHUNKs for the week before physically attending class where they cover additional learning material to increase their understanding.

The MA4027 course was taught by Professor Ralucca Gera and employed two days of self-guided learning, followed by two days of in-class learning. Prof. Gera did not enforce that required CHUNKs had to be completed in order to receive course credit. Instead, users were prompted to select two quizzes to be completed from among the assessment CHUNKlets covered that week. The OS3307 and OS3604 courses, both taught by Professor Michelle Isenhour, provided students with three days of self-guided CHUNK Learning and two days of in-class attendance to discuss the material and complete a group lab activity. In contrast, Prof. Isenhour, did require that students complete CHUNKs to earn credit for the course. We are interested in the differences between each of the courses as a result of instructor guidance. The required CHUNK totals from each course are listed in Table 3.1.

Course	# of CHUNKs Required
MA4027	24
OS3307	44
OS3604	43

Table 3.1: Required number of CHUNKs per course

The table displays a varied number for the required CHUNKs pertaining to each course. We notice that OS3307 required the highest amount, while MA4027 required the least. Due to this variability, we are curious to know the effect produced by a course containing a higher amount of content to keep the interest of the students. The names of each required CHUNK can be found in Appendix A.1. We will look to incorporate this data in Section 3.1.1.

3.1.1 Statistics Data

For our statistical analysis approach, we must first conduct a thorough analysis of the CHUNK Learning data collected on each of the 68 students. The completion of learning content throughout the quarter by students in these three courses is modeled in Figure 3.1.

Through inspection of this data, we can deduce that each class experienced widespread variability in relation to frequency and date completed. For example, MA4027 and OS3604 saw the highest amount of content completed during the beginning of the quarter, whereas OS3307 saw its highest total during the middle of the quarter. This can be attributed to a number of factors such as: due-outs assigned by the instructor or users taking the opportunity to work ahead and complete additional CHUNKs in their course. Also, we note multiple peaks in the data occur on Mondays, while multiple troughs in the data occur during the weekends. This evidence points toward the general trend that users are likely to complete their assignments at the beginning of the academic week rather than in their off time on Saturdays and Sundays.

From the CHUNK Learning user data, we extract two subsets that can be interpreted to measure student performance across all three courses. This tailored data may enable the system to make substantial improvements towards the student learning

Figure 3.1: CHUNK Learning completion report

process that occurs each time a student interacts with a particular CHUNK. We are interested in potentially determining how proficient each student was when engaged in CHUNK Learning for their coursework.

The first subset is the number of CHUNKs completed per user which can be found in Figure 3.2.

We would like to compare a user's number of completed CHUNKs to the number

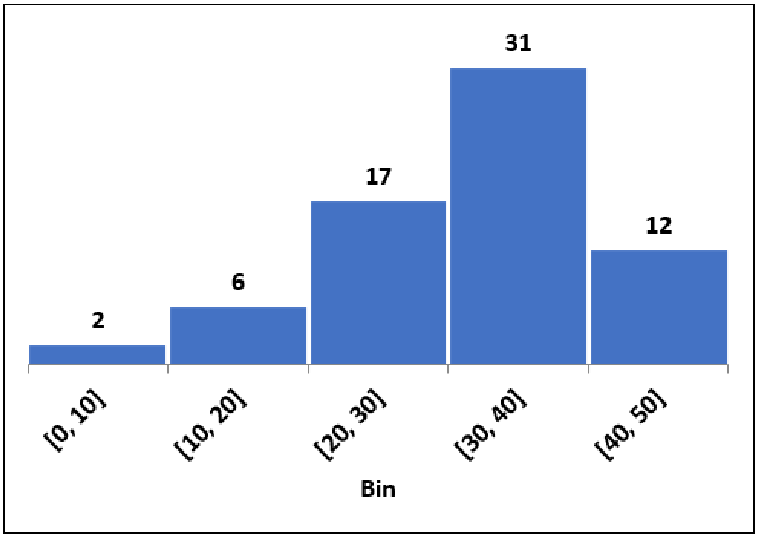


Figure 3.2: Histogram plot of the number of completed CHUNKs among 68 students.

of CHUNKs mandated by their instructors. This information may reveal important trends such as the following: users are satisfied with learning only their assigned course material or users are very interested in learning about fields well outside their current scope of study. Additionally, this information allows us to learn each user's completion rate within the course by observing whether users are completing more than, equal to or less than the required amount of CHUNKs that are assigned to them by their instructors. The histogram plot shows a wide range of the number of CHUNKs being completed by students from each of the three courses. We observe that the [30, 40] bin had the highest frequency with 31 and the [0, 10] bin had the lowest frequency at 2. The data is largely skewed by the fact that each course had a different amount of required CHUNKs, so the MA4027 students would have a much lower average of completed CHUNKs than the students in OS3307 and OS3604. We must account for these variations to effectively utilize the data for a proper statistical analysis.

The second subset of data contains each user's view counts for completed CHUNKs. This value measures the number of times that a user opens a CHUNK with the intention of completing its corresponding learning material. With this information, we are able to potentially gauge a user's interest level in the specified course's learning content. A sample of this data is listed in Table 3.2.

Table 3.2: The number of completed CHUNK views for 10 users

User	Course	# of Completed CHUNK Views
1	MA4027	102
2	MA4027	360
3	MA4027	318
4	OS3307	232
5	OS3307	284
6	OS3307	208
7	OS3604	271
8	OS3604	239
9	OS3604	465
10	OS3604	449

This table represents trends dissimilar to that of Figure 3.2 since there is a wide distribution of completed CHUNK views within each of the course. Although OS3307 contained more required content than MA4027, two users within MA4027 had more content views than three users from OS3307. Many would simply point to the fact that those MA4027 students found the content to be more interesting and valuable than the three OS3307 students. Although this may be the case, we can shift our focus from the student to the learning content itself. As we further inspect the data, we may notice that some CHUNKs are much more interesting than its counterparts. To deduce which CHUNKs these are, we change our mode of analysis to that of network science.

3.1.2 Network Science Data

To properly conduct our network science analysis of the CHUNK Learning data, we must first determine the appropriate relationship to model. Just like in Section 3.1.1, we are interested in the relationship between CHUNK Learning users and the learning content they study. Here, the nodes are of two types: CHUNK Learning users and the CHUNKs from CHUNK Learning. Then each edge represents a connection made between a user and a CHUNK that he or she has watched or completed. We develop a graph that models the various interactions between users and this content.

Once the network is created, We are interested in observing which CHUNKs are most popular and appealing to new students that could potentially be recommended to a user that would like to just sample ideas of a topic before deciding on a deep dive into the topic. So in our case, we are considering the topics that are studied the most and that connect all the CHUNKs of a specific topic. We use the network modeling visualization and analysis program, Gephi, to visualize and run some analysis of such graphs. An example of the completed CHUNK-User network can be found in Figure 3.3.

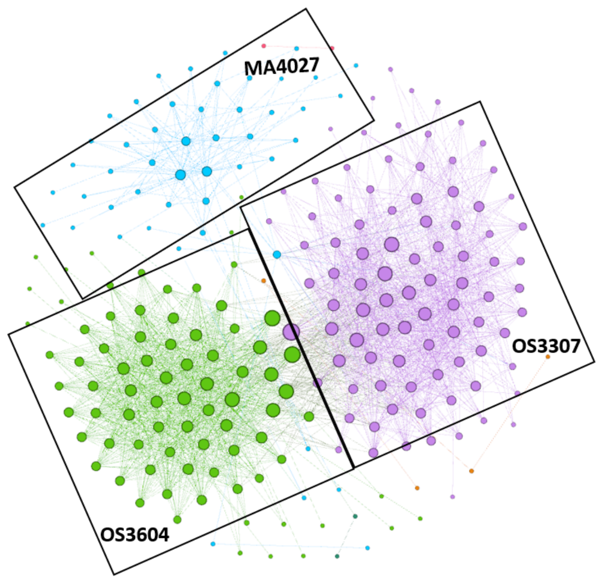


Figure 3.3: Completed CHUNK-User network

We use the following colors to depict CHUNKs, users, and their connections in this network: blue represents nodes and edges related to MA4027, purple represents nodes and edges related to OS3307, and green represents nodes and edges related to OS3604. This network reveals that among the total number of nodes and edges, the majority belong to OS3604 and OS3307. As a result, OS3604 and OS3307 CHUNKs may be more influential than MA4027 CHUNKs. To reduce a course having too much influence as a result of having more nodes, we instead analyze the subgraphs of the network that correspond to CHUNKs and Users from a single course. Therefore, we examine the interactions between users and CHUNKs contained within each of the three courses separately. We use the following colors to represent a type of node: red, blue, cyan, and yellow. A red node represents a user, a blue node represents a required CHUNK, a cyan node represents a recommended CHUNK, and a yellow node represents an optional CHUNK. Also, nodes with a high degree are sized larger than nodes with a low degree. We display the network

We display the CHUNK-User network and its degree distribution for MA4027 in Figure 3.4.

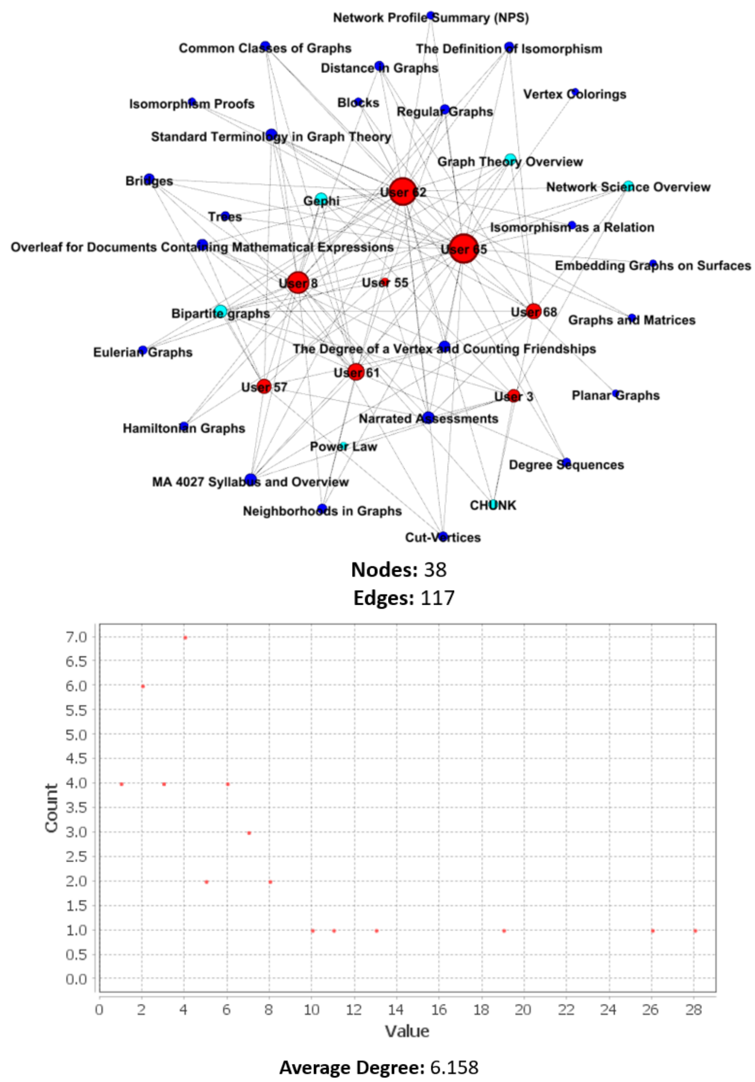


Figure 3.4: MA4027 CHUNK-User network view and degree distribution

This figure shows the 38 nodes are made up of eight students, 24 required CHUNKs and 6 recommended CHUNKs. We notice that the top three nodes by degree are all users and the user nodes are for the most part larger than the CHUNK nodes. This is a result of the fact that there are many more CHUNKs than students.

We now bring attention to Figure 3.5 representing the users and CHUNKs of the OS3307 course.

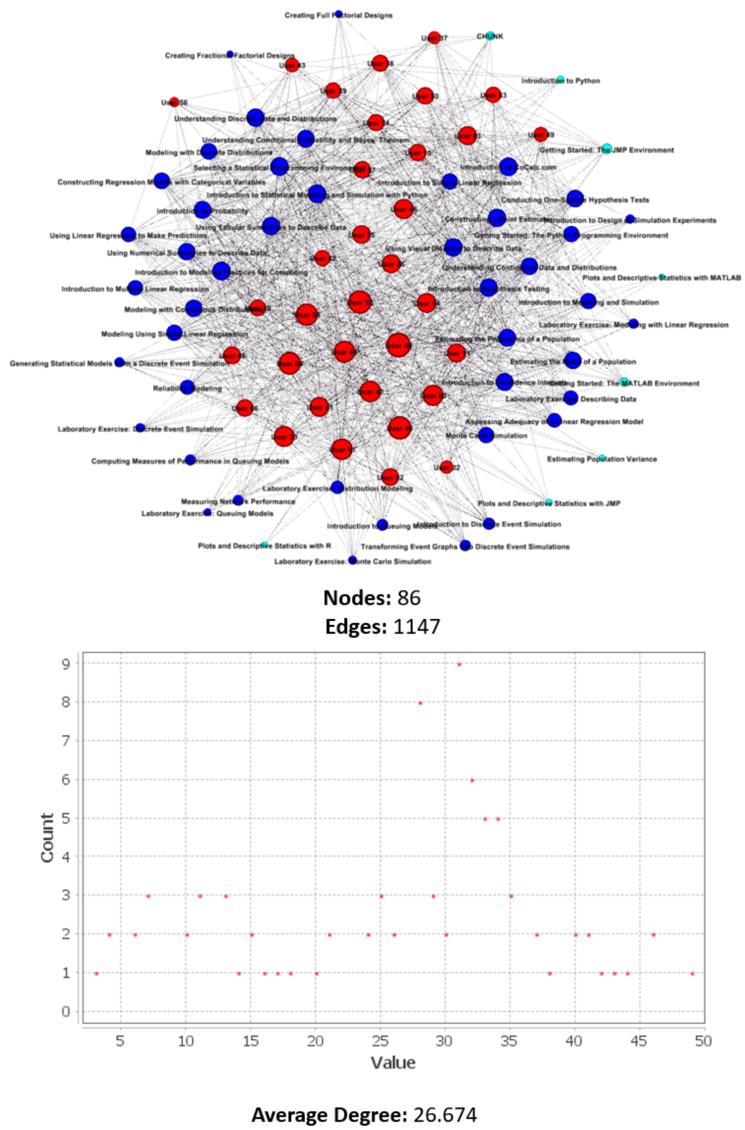


Figure 3.5: OS3307 CHUNK-User network view and degree distribution

OS3307 network's 86 nodes are made up of 34 students, 55 required CHUNKs and 8 recommended CHUNKs. In this network, the student nodes continue to have the highest degree, though there are several required CHUNKs that are similar in degree to some of the students. We can explain this trend due to the higher number of students in the class, which makes the ratio of students to CHUNKs more evenly balanced.

We present the nodes and edges of the OS3604 network in Figure 3.6.

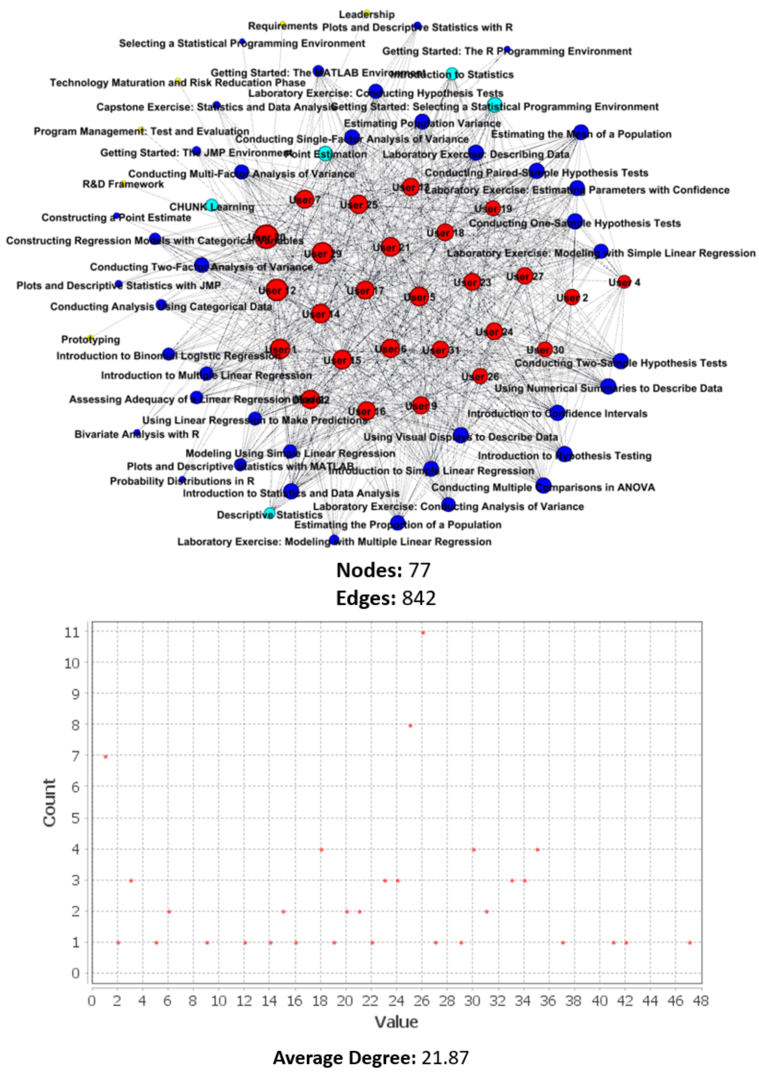


Figure 3.6: OS3604 CHUNK-User network view and degree distribution

This network displays 77 nodes made up of 26 students, 40 required CHUNKs, 5 recommended CHUNKs and 6 optional CHUNKs. The network and degree distribution for OS3604 both look very similar to that of OS3307 since they possess a similar amount of students and CHUNKs.

We would like to explore which CHUNKs were the most important within each topic based on key concepts such as k-core and betweenness centrality. Moreover, we are curious to see if these parameters are more likely to reveal required CHUNKs than the other types. Further, we look to determine whether learning content was prioritized during a certain period in the course, whether early on or later towards the final examinations.

3.1.3 Limitations

Though the CHUNK Learning data used in this project has enormous potential in developing important solutions that can be implemented in other AESs as well, there are certain limitations that must be addressed. First, we must consider the background of the users analyzed in this study. All 68 users attend NPS, a graduate school with both homogeneous and heterogeneous characteristics. Its student population mostly consists of military officers from the U.S. Navy and other U.S. military branches, but there are also international students who are military officers in their home countries. These users are also enrolled in either an Applied Mathematics or Operations Research course, both in the science, technology, engineering, and mathematics (STEM) field of study, so they may have very similar learning approaches.

The second limitation is that the data is potentially affected by human error that occurs when users navigate in CHUNK Learning. Here, we present two different types of human error specific to CHUNK Learning. The first type, or Category 1, is an error of commission. Applied to our situation, errors of commission happen when a user clicks complete on an Activity (Figure 2.2) even though he/she did not complete the activity. As a result, the user would be credited with a high number of completed CHUNKs but have a much lower number of completed CHUNKs views. This case occurs when users engage in negligent behavior and simply aim to get credit for learning with minimal effort. Unfortunately, the current CHUNK Learning system, which is still a prototype, does not currently contain any controls, such as mandatory time limits or click tracking, to prevent such behaviors. The second type of error, or Category 2, is an error of omission. Errors of omission occur when a user simply forgets to click the complete checkbox after completing an Activity. Therefore, a user would have a low number of completed CHUNKs but, potentially, a high number of completed CHUNK views. This is a knowledge-base error and can occur as a result of the user being unaware that they need to perform additional actions to complete an activity, in which case the instructor may have not properly instructed the user on how to use the system.

3.2 Methodology Overview

This section outlines the methods we perform to conduct a statistical and network science analysis of user engagement within CHUNK Learning. In Section 3.2.1, we discuss the statistical analysis approach to user analytics within CHUNK. Further, we discuss the construction of the algorithm that is aimed to equip the instructor with deeper knowledge of their students, ultimately allowing them to make major improvements to their learning paths. Then in Section 3.2.2, we introduce the network science analysis methodology that provides deeper insights from the user data available to us such as the identification of the most important CHUNKs in a course.

3.2.1 Statistical Analysis Approach

We use this tailored data to identify key trends that exist within each class of users. In order to proceed, we must determine the value of two specific variables: the completed CHUNKs/required CHUNKs (CR) score and the views of completed CHUNKs/completed CHUNKs (VC) score. This data allows us to solve for the two variables that the project's main methodology is based on. We know that CHUNK is competency-based learning, but we can still use analytics to determine user proficiency within a CHUNK. We identify the standards as: does not meet the standard, partial meets the standard, meets the standard, and exceeds the standard. These standards can be represented as Level 1, Level 2, Level 3, Level 4 respectively.

We now introduce a few definitions that will help with our analysis, by creating ways of measuring completed content as a fraction of required content, and representing the number of times a user views content he/she completed as a fraction of completed content. We compute these ratios based on the number of completed CHUNKs. We may be able to unearth varying patterns of user activity that exist within CHUNK Learning so that we can better understand how to improve each user's experience. First, we start by defining CR, which enables us to examine whether or not users are completing their required CHUNKs as prescribed by their instructors.

Definition 3.2.1 *Let C_i be the amount of content (i.e., CHUNK) completed per user i, and R_j be the amount of content required for course j. We define CR_i as:*

$$CR_i = \frac{C_i}{R_j}$$

CR_i provides a value that represents completion rate or the extent to which a user completed more than, equal to or less than the amount of content required for course j. This value, although useful, is not standardized across the different courses and requires an additional step to produce a normalized score for each user in every course. Hence, we compute a z-score for each CR_i, as referenced in Definition 2.3.3.

Definition 3.2.2 *Let CR_i be the amount of completed content divided by the amount of content required in course j per user i, $\overline{CR}_j$ be the average CR among the entire population in course j, and σ_j be the standard deviation for the CR in course j. We define Z_{CR} for each user i as:*

$$Z_{CR} = \frac{CR_i - \overline{CR}_j}{\sigma_j}$$

We use this z-score to standardize CR, and we now represent it as the user's content CR score. Each user has a CR score corresponding to completion rates for CHUNKs. We aim to use these scores to indicate one component of competence based purely on a user's completion rate.

Next, we introduce the concept of VC, which provides the foundation for calculating a user's interest level.

Definition 3.2.3 *Let V_i be the number of times user i viewed the content he/she completed, and C_i be the amount of content completed by user i. We define the VC_i*

score as:

$$VC_i = \frac{V_i}{C_i}$$

VC_i provides a ratio of completed content views to completed content for each user i. This value, like CR_i, must be standardized to account for differences between each of the courses and so we also compute a z-score for VC_i, which we represent as Z_{VC}.

Definition 3.2.4 *Let VC_i be the ratio of completed content views to completed content for each user i in course j, $\overline{VC}_j$ be the average VC among the entire population in course j, and σ_j be the standard deviation for the VC in course j. We define the Z_{VC} for user i as:*

$$Z_{VC} = \frac{VC_i - \overline{VC}_j}{\sigma_j}$$

We again use the z-score to standardize the VC and introduce Z_{VC} as the VC score, a method to measure the learning interest of a user. We calculate CHUNK VC scores for each user enrolled in a course.

We look to incorporate the CR and VC scores into an algorithm to clearly identify learners by their competence level. We coin this algorithm the standardized algorithm of required CHUNKs (STAR). STAR measures the overall competence level of each student within their class by their STAR score. Each user's CR and VC z-scores will be assigned to one of four ratings: very high (VH), high (H), low (L) and very low (VL). The criteria for these ratings are as follows:

1. Very High: $Z \geq 1$
2. High: $1 > Z > 0$
3. Low: $0 < Z < -1$
4. Very Low: $Z \leq -1$

With these ratings, we are able to classify the whole range of student scores. We obtain each user's STAR score using a point system, where both the CR and VC scores have an assigned point total. We accomplish this by allocating points to each of the four ratings in the following manner:

1. Very High: 3 points
2. High: 2 points
3. Low: 1 points
4. Very Low: 0 points

The creation of these four ratings allows for users to fall under one of 16 different combinations that can be used to describe the possible types of users that exist based upon CR and VC scores. Furthermore, we can assign point values to each of these groups and determine a manner in which to assign each one into competence levels ranging from Level 1 to Level 4. We introduce this as the STAR Rating System. We list and describe these 16 groups in Table 3.3.

Table 3.3: Description of the STAR Rating System groups

Groups	**Description**
VL/VL	user completes few required CHUNKs and barely views the content
VL/L	user completes few required CHUNKs and views some content
L/VL	user completes some required CHUNKs and barely views the content
L/L	user completes some required CHUNKs and views some content
VL/H	user completes few required CHUNKs and views most of the content
H/VL	user completes most of the required CHUNKs, but barely views the content
VL/VH	user completes few required CHUNKs, but views all of the content
L/H	user completes some required CHUNKs and views most of the content
H/L	user completes most of the required CHUNKs and views some content
VH/VL	user completes all required CHUNKs, but barely views the content
L/VH	user completes some required CHUNKs, views all of the content
VH/L	user completes all required CHUNKs and views some content
H/H	user completes most of the required CHUNKs and views most of the content
H/VH	user completes most of the required CHUNKs and views all of the content
VH/H	user completes all required CHUNKs and views most of the content
VH/VH	user completes all required CHUNKs and views all of the content

We can use these 16 combinations to represent the variability in learning experienced among the students in MA4027, OS3307 and OS3604. These groups also unlock potentially critical trends that may be specific to a certain course or its instructor. Moreover, we are able to assign STAR points to these groups, which provide the metric for determining a user's competence level. The STAR Rating System is shown in Table 3.4.

Table 3.4: The STAR Rating System includes 16 different user groups based on CR and VC scores

Groups	Points	Competency Level
VL/VL	0	1
VL/L	1	1
L/VL	1	1
L/L	2	2
VL/H	2	2
H/VL	2	2*
VL/VH	3	2**
L/H	3	2**
H/L	3	3*
VH/VL	3	3*
L/VH	4	3**
VH/L	4	3
H/H	4	3
H/VH	5	4
VH/H	5	4
VH/VH	6	4

* Category 1 Human Error
** Category 2 Human Error

We are specifically interested in the VH/VL and VL/VH combinations, which highlights students whom display extreme polarity regarding CR and VC, and are at the most risk for Category 1 human error and Category 2 human error respectively. There are also four other user groups which may have the potential for human error. We note that the L/H and L/VH groups are at low-risk for Category 1 human error, whereas the H/VL and H/L groups are at low-risk for Category 2 human error.

Our statistical analysis of the combined CR and VC scores will use the following methodology:

STAR: a CHUNK Learning User Competency Algorithm

1. First, extract CHUNK Learning data for a user i.
2. Solve for CR_i as referenced in Definition 3.2.1 .
3. Use the CR_i to find the user's Z_{CR} as explained in Definition 3.2.2.

4. Assign rating #1 to user based on their CR score.
5. Compute VC_i as referenced in Definition 3.2.3.
6. Use VC_i to determine the user's Z_{VC}, as outlined in Definition 3.2.4.
7. Assign rating #2 to user based on their CR score.
8. Use both ratings to determine user group and competence level according to the STAR Rating System in Table 3.4.

3.2.2 Network Science Analysis Approach

For this thesis, we aim to identify the most important subgroup within the network through the use of network analysis within CHUNK Learning. Specifically, we focus on determining the core of each course's network and the centrality of CHUNKs within each course. We look to conduct a k-core analysis to segregate the network's core from its periphery. Then, we conduct an in-depth analysis of betweenness centrality and eigenvector centrality among all course CHUNKs.

In Gephi, we can learn many important details of a network by applying filters which aim to remove certain types of nodes. For this project, we apply the k-core filter as seen in Figure 3.7.

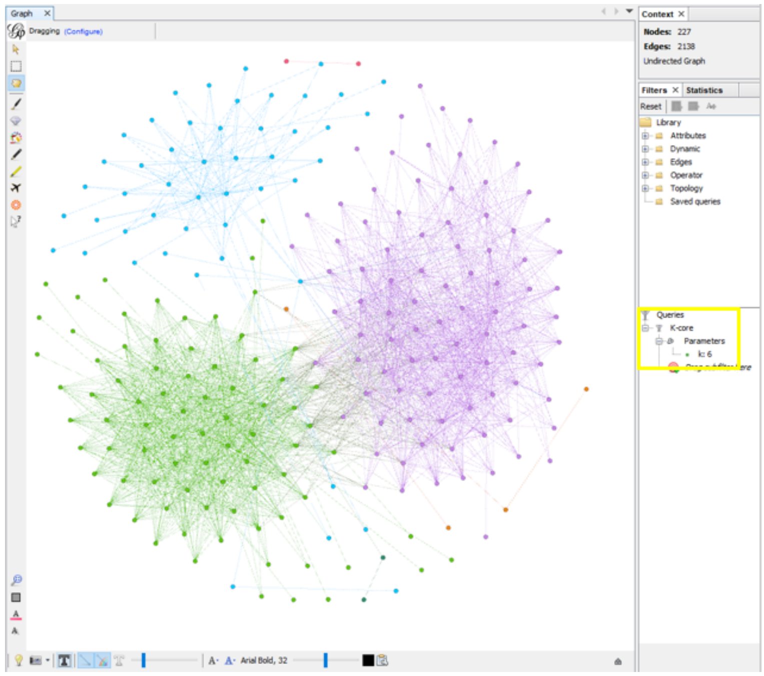
By removing each of the lower order k-cores, we are able to truly segregate the network's core for each of the three courses. This core may reveal the set of CHUNKs that are most influential based on students' patterns of accessing the content. At this point, we introduce the following methodology to determine a network's core in Gephi.

Identifying the Core of a Network:

1. In Gephi, import each course's completed CHUNK data to obtain a network.
2. Use k-core filter within the Topology folder to filter lower order k-cores.
3. Increase the number in k-core settings until the network is completely empty.
4. Decrease the number by the value 1 to reveal the network's core.

Moreover, we are interested particularly in the amount of influence that each node has, specifically the influence of a single CHUNK. In order to discover this data, we will analyze betweenness centrality and eigenvector centrality, both indicators of

Figure 3.7: k-core filter



importance, within CHUNK Learning. We display the calculations we will use in Figure 3.8.

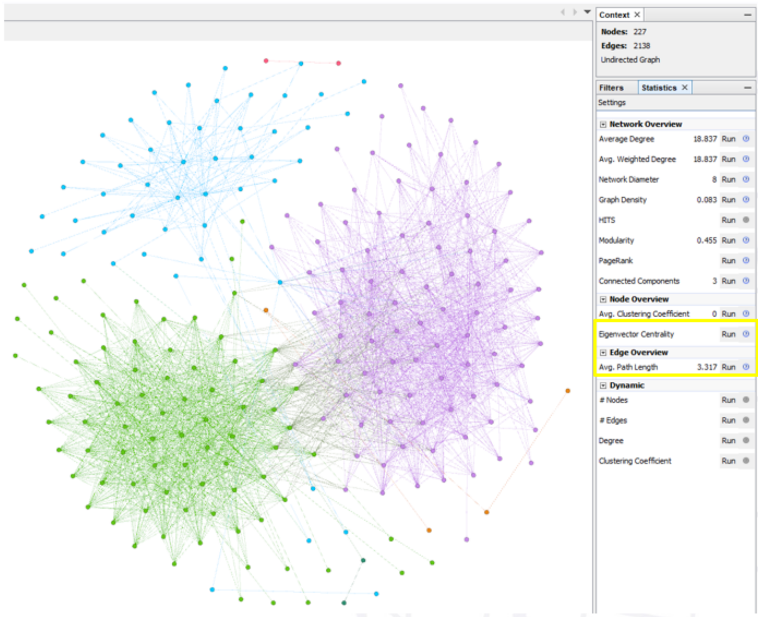
Next, we discuss how we are able to use the internal algorithms within Gephi to determine centrality values for our networks.

Determining the Centrality of a Network:

1. Import each course's completed CHUNK data into Gephi.
2. Run Average Path Length and Eigenvector Centrality calculations under the Statistics tab.
3. Switch to the Data Laboratory, and observe the betweenness centrality and eigenvector centrality values for each of the nodes.

We seek to gain insight on which CHUNK may be of most value within its network. With the use of these tools, we analyze the CHUNK Learning system from a different lens. Gephi allows us to visualize how well-connected each CHUNK is within its network and we are able to conduct valuable analysis on the strength of these

Figure 3.8: Centrality algorithm in Gephi



connections. This data can be easily utilized by the system to develop an abridged version of a course that may give explorers a highly efficient and effective first look into whether a course may be right for them.

This chapter presented this project's CHUNK Learning data and the various mathematical methodologies we implemented for our analysis. In the next chapter, we display the results of the STAR algorithm to determine CR score, VC score, STAR ratings, STAR groups and STAR levels for each user. Then, we describe the results of the network science analysis in Gephi to discover the most important CHUNKs in MA4027, OS3307 and OS3604.

CHAPTER 4:
Results and Analysis

The following chapter discusses the results of the two analysis methods used in this thesis. Section 4.1 examines the results of the statistical analysis using the STAR Algorithm on our CHUNK Learning user data. Section 4.2 presents the results of the network science algorithm using Gephi. Moreover, we identify whether key trends exist in each analysis that can provide deeper insight into user engagement within CHUNK Learning.

4.1 Statistical Analysis Results

We break down the statistical analysis for MA4027, OS3307 and OS3604 into three parts: (1) CR and VC scores, (2) STAR groups, and (3) STAR competence levels. We display the results for MA4027 in Section 4.1.1, OS3307 in Section 4.1.2 and OS3604 in Section 4.1.3.

4.1.1 MA4027 Results and Analysis

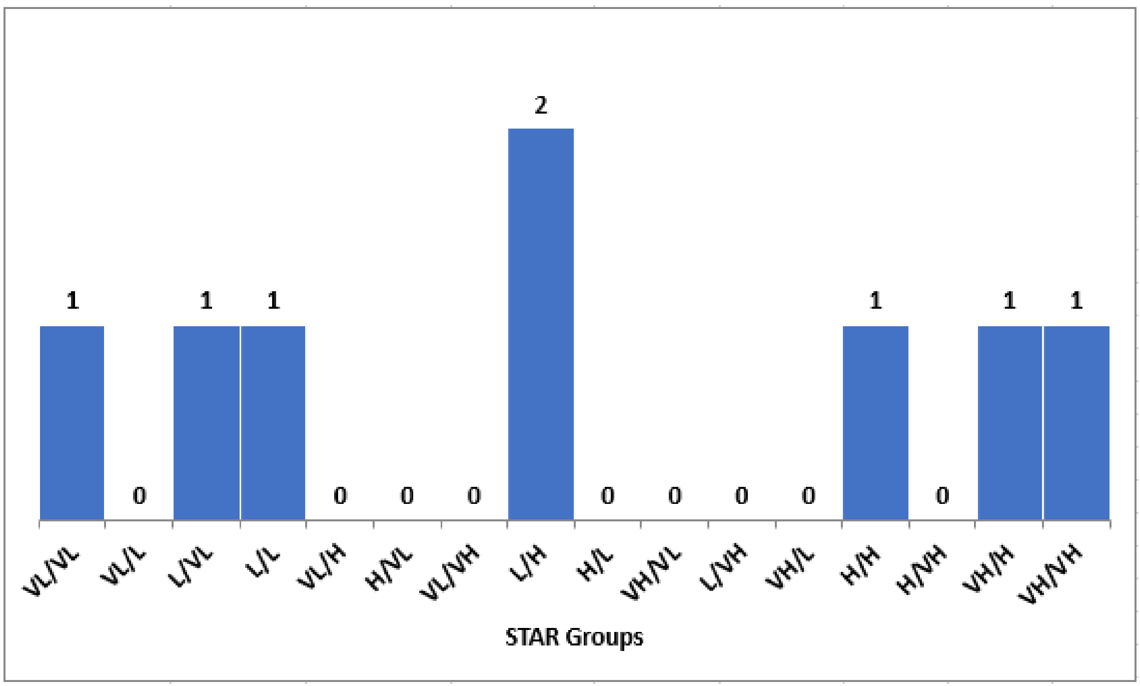
This section presents the results and analysis of the Statistical Analysis methodology applied to the eight students in the MA4027 course. First, we show the CR and VC scores with their corresponding ratings in Table 4.1.

Table 4.1: MA4027 Z_{CR} and Z_{VC} scores and ratings

User ID	Z_{CR} **Score**	Z_{CR} **Rating**	Z_{VC} **Score**	Z_{VC} **Rating**
User 3	-0.66	L	0.76	H
User 8	0.66	H	0.92	H
User 55	-1.38	VL	-1.52	VL
User 57	-0.56	L	-1.21	VL
User 61	-0.25	L	-0.56	L
User 62	1.07	VH	1.16	VH
User 65	1.58	VH	0.13	H
User 68	-0.46	L	0.32	H

Among the eight students, the Z_{CR} scores range from -1.38 to 1.58, while the Z_{VC} scores range from -1.52 to 1.16. There is not much disparity between the minimum and maximum scores which may be a result of the relatively small number of students in the class. We also notice that the students fared worse in Z_{CR} with five students having a low or very low rating but fared better in Z_{VC} with five students scoring a high rating or better. This trend indicates that most of the students did not prioritize completion of the required CHUNKs, which is consistent with Prof. Gera's instruction stating that required CHUNKs completion was not necessary to obtain course credit. Next, we display the breakdown of users by STAR group in Figure 4.1.

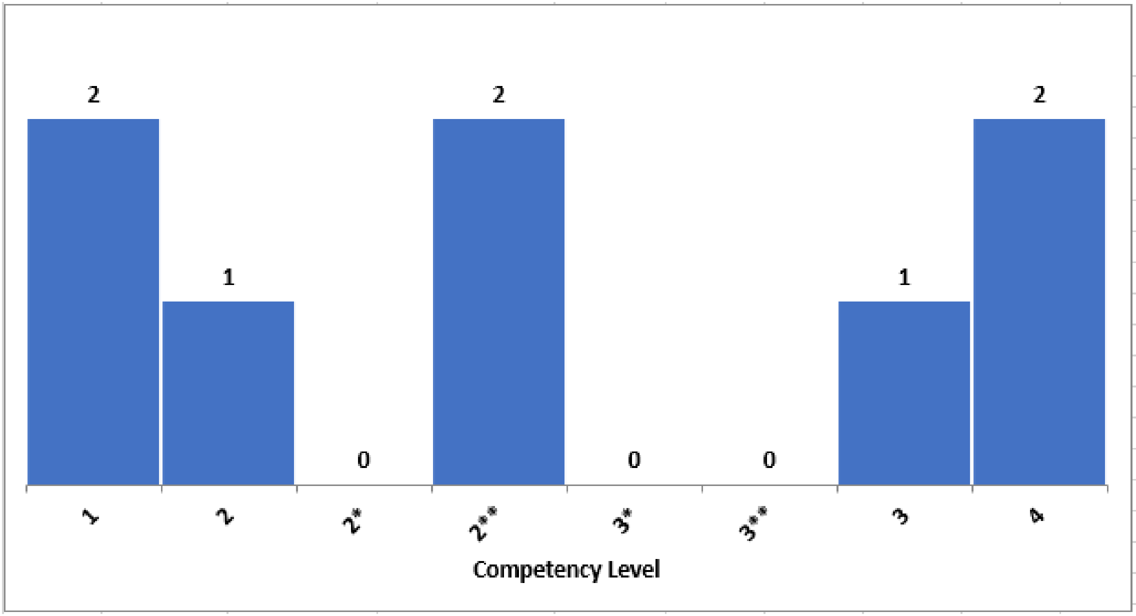
Figure 4.1: MA4027 STAR analysis



In Figure 4.1, there are seven different groups represented by the eight MA4027 students, which indicates a wide array of CHUNK Learning experiences. Within the course, we note that one student represented the best group, VH/VH, and one student represented the worst group, VL/VL. In addition, there were two students in the L/H category which we identify as cases of low-risk Category 2 human error. These cases are further highlighted in Figure 4.2.

In Figure 4.2, we observe the distribution of students by STAR level. It is evident

Figure 4.2: MA4027 STAR levels



* Category 1 Human Error, ** Category 2 Human Error

that three students performed well and exceeded the standard whereas the other five students fell into a STAR Level 2 or lower. We can attribute the high percentage of students below the standard to the preponderance of low Z_{CR} scores. With this data, Prof. Gera can easily prioritize her time to these five students, starting with the two students placed in STAR Level 1 who do not meet the standard. Moreover, she can address any existing issues concerning the two students that were at low risk for Category 2 human error. In the analysis of this small class size, the STAR system easily recognizes students that may need additional assistance inside CHUNK Learning.

4.1.2 OS3307 Results and Analysis

In this section, we display the results and analysis of the Statistical Analysis methodology on the 34 students within the OS3307 course. First, we show a portion of the CR and VC scores with their ratings in Table 4.2.

The Z_{CR} analysis shows a much larger range of scores from -2.52 to 1.91 compared to MA4027 since there are many more students. We see increased variability due to

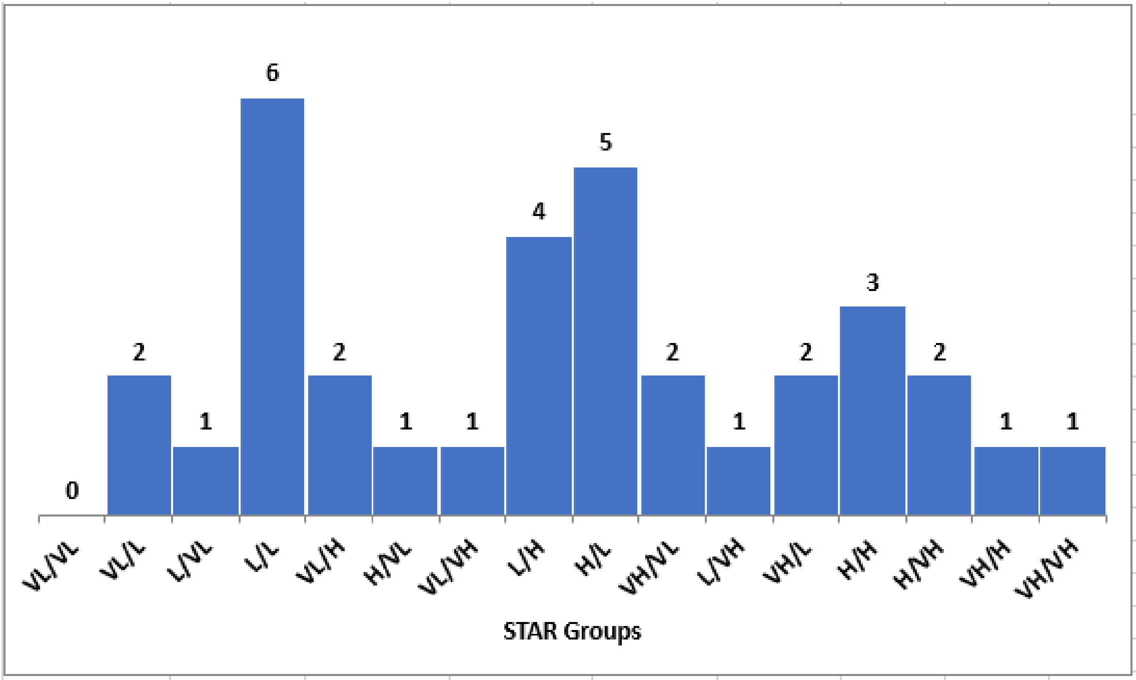
Table 4.2: Sample of OS3307 Z_{CR} and Z_{VC} scores and ratings

User ID	Z_{CR} **Score**	Z_{CR} **Rating**	Z_{VC} **Score**	Z_{VC} **Rating**
User 28	1.56	VH	-1.09	VL
User 32	1.09	VH	-1.36	VL
User 33	1.91	VH	-0.09	L
User 34	1.09	VH	-0.89	L
User 45	1.91	VH	3.02	VH
User 49	1.09	VH	0.28	H
User 56	-1.01	VL	-0.40	L
User 58	-1.12	VL	0.13	H
User 59	-1.59	VL	1.19	VH
User 60	-1.59	VL	-0.42	L
User 66	-2.52	VL	0.28	H

11 students earning a VH or VL rating, 32.3% of the class. OS3307's Z_{VC} scores also experienced higher variability with scores ranging from -1.36 to 3.02 and nine students scoring a VH or VL rating. We point out one specific user, User 45, who had a profound impact on the rest of the data as a VH/VH user with a $Z_{CR} = 1.91$ and $Z_{VC} = 3.02$. As a result of this extremely high Z_{VC} score, we determined that 19 out of 34 students received a low or very low rating in Z_{VC}. We recall that Prof. Isenhour mandated the completion of required CHUNKs in the course which should have resulted in a much tighter distribution with Z_{CR} scores closer to the sample mean of 0. Instead, students like User 45 with extremely high z-scores can negatively affect the rest of the z-scores in the course. We now display the STAR analysis in Figure 4.3.

In Figure 4.3, we observe that users were placed into 15 out of the 16 possible STAR combinations, with zero users in the VL/VL group. The effect of User 45 is even further evident since there is one VH/VH user and no VL/VL users, unlike the distribution in MA4027. By inspection, we notice two users in the VH/VL group, indicative of high-risk for Category 1 human error, and one user in the VL/VH group, indicative of high-risk for Category 2 human error. Prof. Isenhour may be able to now prioritize these students as high-risk for human error through the use of the STAR algorithm. Users in these groups may easily see their competence levels raise by resolving these

Figure 4.3: OS3307 STAR analysis



errors. We now present the competence levels of OS3307 in Figure 4.4.

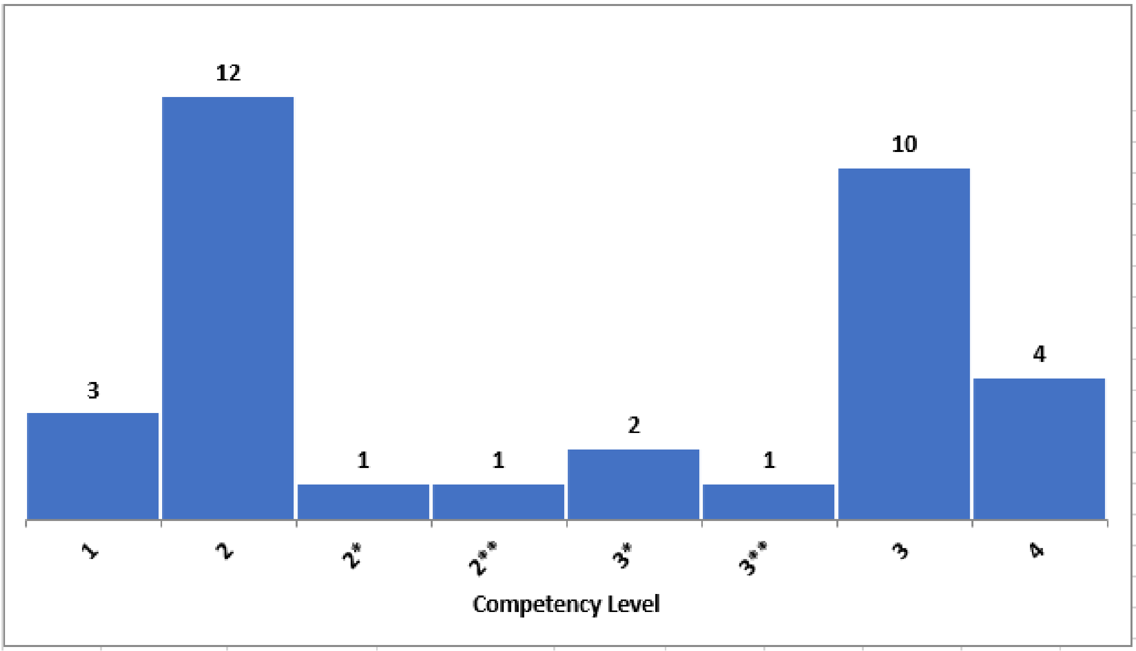
When we breakdown student scores into STAR levels, we notice a similar distribution to that MA4027. We observe 15 students in Levels 1 and 2, earning a partially meets the standard rating or lower, compared to 14 students in Levels 3 and 4. The remaining five students fell into levels potentially affected by Category 1 and Category 2 human error. So, the STAR algorithm reveals a total of 20 students that may need additional instruction with the 3 students in Level 1 requiring the most attention. Prof. Isenhour may be able to utilize the STAR system to effectively determine which students to focus her time towards.

4.1.3 OS3604 Results and Analysis

This section presents the results and analysis of the Statistical Analysis methodology applied to the 24 students in the OS3604 course. First, we show a sample of the CR and VC scores with their corresponding ratings in Table 4.3.

The Z_{CR} analysis for OS3604 displays a similar range of scores to that of OS3307

Figure 4.4: OS3307 STAR levels



* Category 1 Human Error, ** Category 2 Human Error

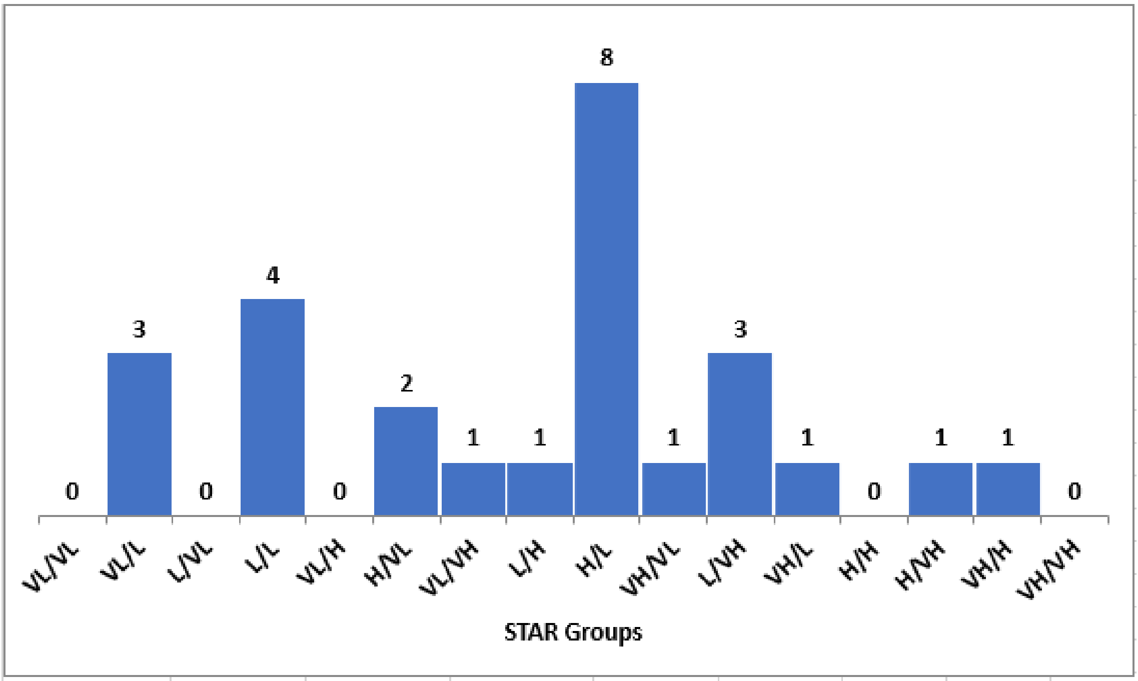
Table 4.3: Sample of OS3604 Z_{CR} and Z_{VC} scores and ratings

User ID	Z_{CR} **Score**	Z_{CR} **Class**	Z_{VC} **Score**	Z_{VC} **Class**
User 12	1.81	VH	0.40	H
User 19	-1.30	VL	-0.28	L
User 20	2.50	VH	-0.15	L
User 26	-1.13	VL	-0.64	L
User 29	1.46	VH	-1.11	VL
User 30	-1.30	VL	-0.08	L
User 4	-2.17	VL	1.61	VH

with scores from -2.17 to 2.50 but shows less dispersion. We note there were only seven students with a VH or VL rating, 26.9% of the total class. Though, OS3604's Z_{VC} scores showed little variation in comparison to OS3307. Unlike in the previous courses, there were no students with a VH/VH or VL/VL rating. We introduce the STAR analysis for OS3604 in Figure 4.5.

Through inspection of the STAR group distribution in Figure 4.5, there is significant

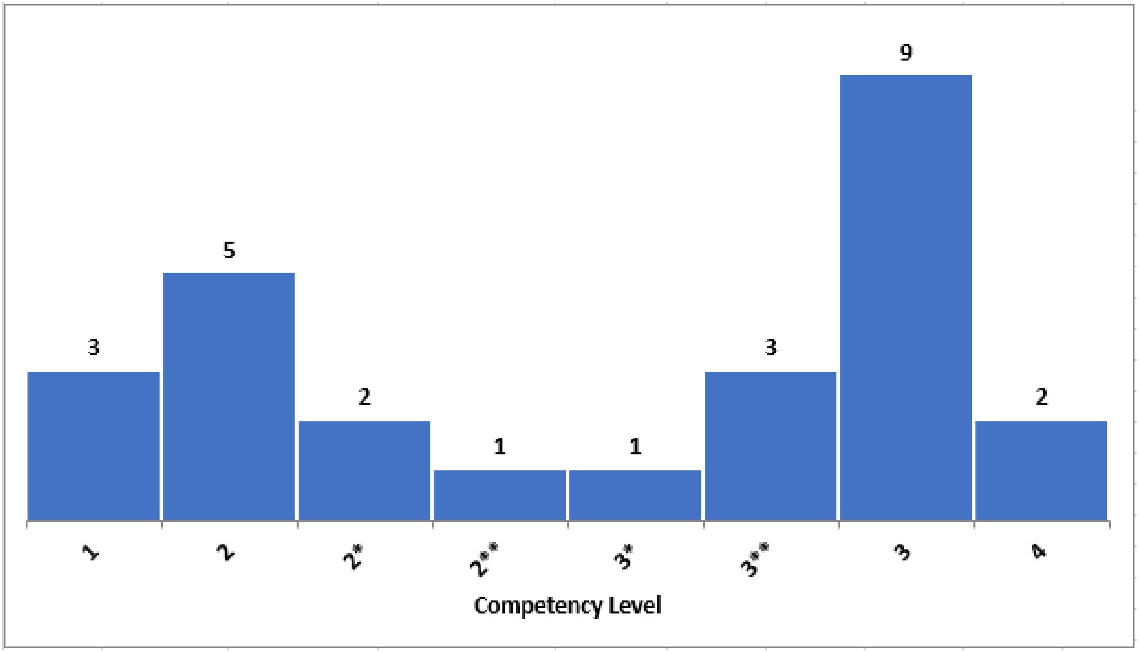
Figure 4.5: OS3604 STAR analysis



concentration in the middle of the graph. This trend indicates that the majority of students in OS3604 scored closer to the mean scores for both Z_{CR} and Z_{VC}. The distribution was positively affected by the high number of students in the H/L category, the largest STAR group in the class. We can deduce that the students in OS3604 were generally more motivated to complete course content and more interested in learning course content than students in MA4027 and OS3307. We can correlate this positive effect to Prof. Isenhour's directions to complete required CHUNKs. Next, we present the STAR levels of OS3604 in Figure 4.6.

The STAR level breakdown for OS3604 reveals that the majority of students, 15 out of 26, met the standard and achieved a STAR level of 3 or higher. While OS3604 had more students with possible human error than OS3307, it had a much lower number of students at a STAR level of 2 or lower. Based on these results, Prof. Isenhour may be able to determine key differences between her two courses that allowed the students in OS3604 to thrive. Just like in OS3307, Prof. Isenhour would isolate the three students with a STAR level of 1 to pinpoint the direct cause of their low scores.

Figure 4.6: OS3604 STAR levels



* Category 1 Human Error, ** Category 2 Human Error

4.2 Network Science Analysis Results

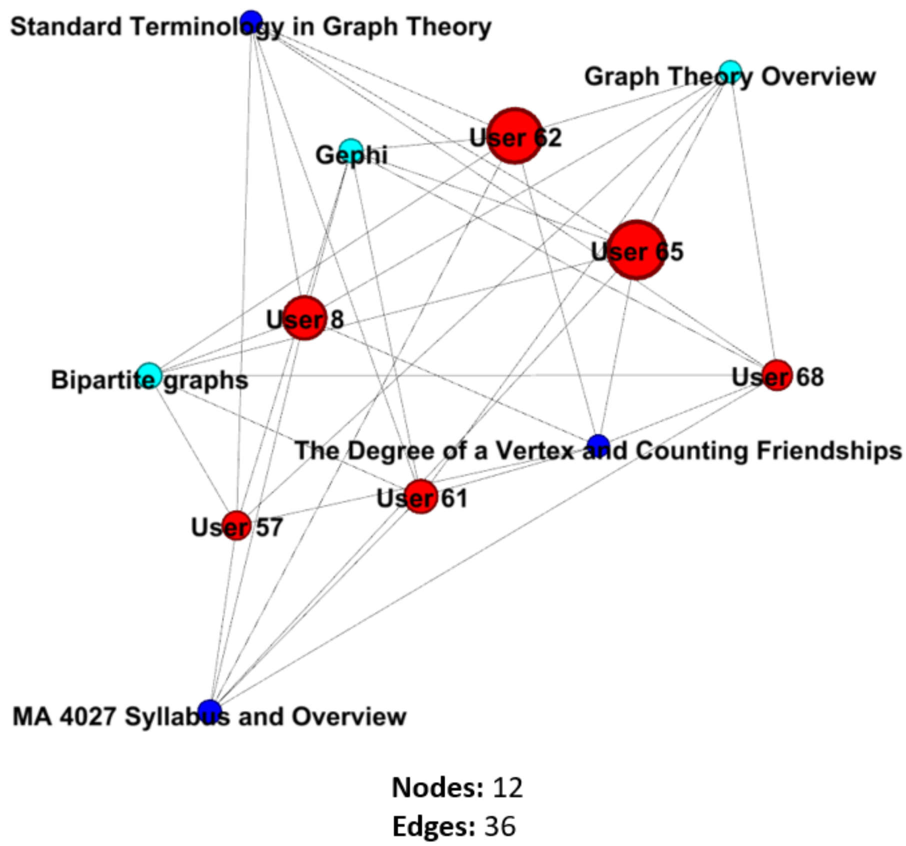
For the Network Science analysis of MA4027, OS3307 and OS3604, we focus on the two following parameters: (1) k-core and (2) betweenness and eigenvector centrality. We explain the results and analysis of MA4027 in Section 4.2.1, OS3307 in Section 4.2.2 and OS3604 in Section 4.2.3.

4.2.1 MA4027 Results and Analysis

In this section, we outline the results of the network science analysis conducted on the CHUNK-User MA4027 Network. First, we display the results of the k-core analysis in Figure 4.7.

We reduced the original MA4027 network of 38 nodes and 117 edges to a 6-core network of 12 nodes and 36 edges. In terms of nodes, we are left with 3 required CHUNKs and 3 recommended CHUNKs. All of these CHUNKs were presented in the first two weeks of the course. Thus, we may deduce that user activity was highest during the onset of the course as some users declined in activity as the course

Figure 4.7: MA4027 k-core analysis



progressed. Moreover, these CHUNKs may have been accessed more throughout the course since they were elementary and representative topics, a goal that we have for this research.

Next, we list the top CHUNKs for each centrality measure in Tables 4.4 and 4.5, where CHUNKs are highlighted in yellow if they appear in both rankings.

Through inspection, we clearly see the centrality results are quite similar for both betweenness centrality and eigenvector centrality. It is important to notice the overlap between the k-core and the centralities results. Moreover, centralities bring to light more topics. By measuring the centrality values of the MA4027 CHUNKs, we are introduced to two additional CHUNKs, Narrated Assessments and Overleaf, that were not previously included within the k-core analysis. Thus, Prof. Gera can efficiently choose among these eight CHUNKs to potentially offer a half-course or smaller course

Table 4.4: Top 5 MA4027 CHUNKs based on betweenness centrality

CHUNK Name	Betweenness Centrality
Bipartite Graphs	35.61
Narrated Assessments	29.60
MA4027 Syllabus and Overview	16.86
Gephi	16.86
Overleaf	13.49

Table 4.5: Top 5 MA4027 CHUNKs based on eigenvector centrality

CHUNK Name	Eigenvector Centrality
Bipartite Graphs	0.52
Gephi	0.5
MA4027 Syllabus and Overview	0.5
The Degree of a Vertex and Counting Friendships	0.47
Standard Terminology in Graph Theory	0.47

offering to exploratory students.

4.2.2 OS3307 Results and Analysis

In this section, we discuss the results of the network science analysis conducted on the CHUNK-User OS3307 Network. We now introduce the results of the k-core analysis in Figure 4.8.

The k-core analysis revealed a 23-core network for OS3307. The core eliminated 26 nodes and 309 edges from the network. There are 60 total nodes, composed of 30 students and 30 required CHUNKs.

After analyzing the CHUNKs in the core, we observe that the chosen CHUNKs had the highest degree among the entire group and were introduced throughout the course as opposed to the first few weeks as in MA4027. It is important to point out that no recommended CHUNKs are part of the core, further proving that required CHUNKs possess more value than recommended CHUNKs.

Next, we discuss the centrality analysis for OS3307 in Tables 4.6 and 4.7.

Figure 4.8: OS3307 k-core analysis

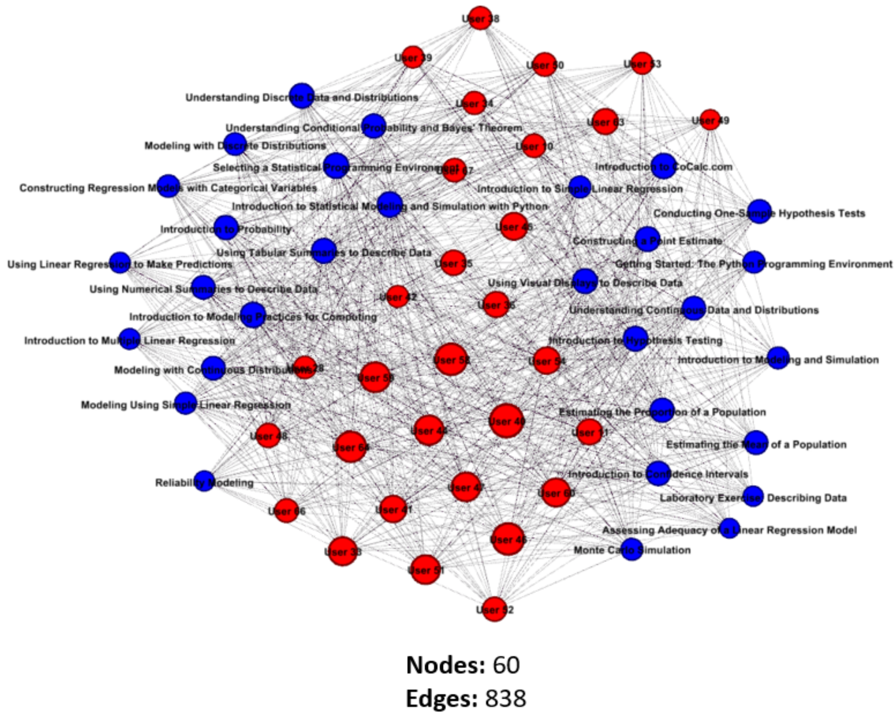


Table 4.6: Top 5 OS3307 CHUNKs based on betweenness centrality

CHUNK Name	Betweenness Centrality
Selecting a Statistical Programming Environment	48.71
Introduction to Modeling Practices for Computing	48.71
Introduction to Stat. Modeling and Sim. with Python	48.71
Introduction to CoCalc.com	48.71
Understanding Discrete Data and Distributions	46.09

The tables clearly indicate four required CHUNKs that appear in both rankings for betweenness centrality and eigenvector centrality. Each of the rankings had a different fifth CHUNK which is a result of the two contrasting approaches towards centrality. Unlike in MA4027, each of these CHUNKs were found in the network's core analysis. Therefore, Prof. Isenhour would be able to effectively build a smaller course using each of these seven CHUNKs.

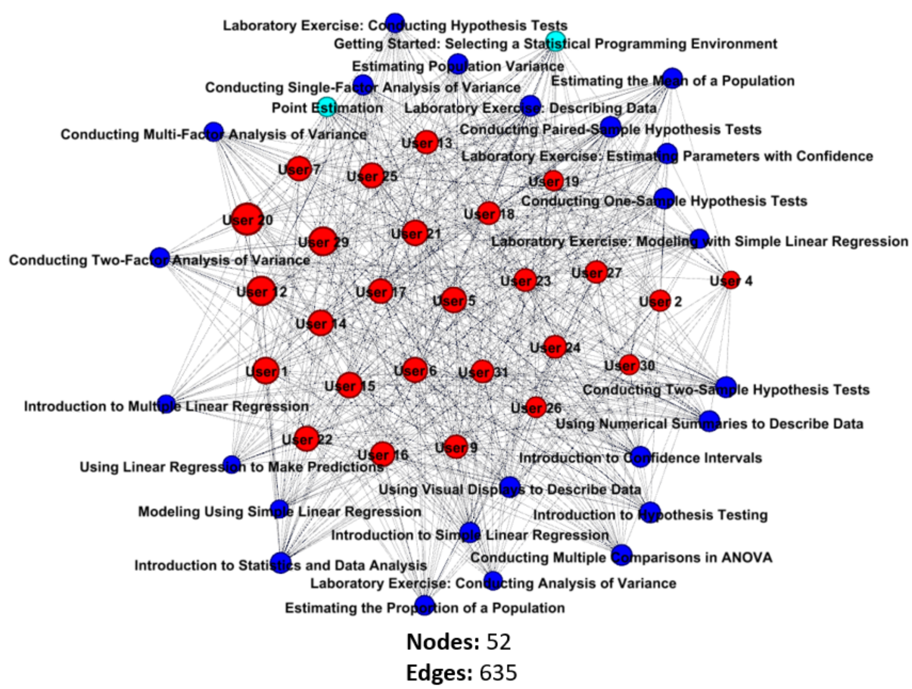
Table 4.7: Top 5 OS3307 CHUNKs based on eigenvector centrality

CHUNK Name	Betweenness Centrality
Introduction to Modeling Practices for Computing	0.88
Selecting a Statistical Programming Environment	0.88
Introduction to Stat. Modeling and Sim. with Python	0.88
Introduction to CoCalc.com	0.88
Introduction to Hypothesis Testing	0.87

4.2.3 OS3604 Results and Analysis

In this section, we outline the results of the network science analysis conducted on the CHUNK-User OS3604 Network. We now introduce the results of the k-core analysis in Figure 4.9.

Figure 4.9: OS3604 k-core analysis



We reduced the original OS3604 network of 77 nodes and 842 edges to a 20-core network of 52 nodes and 635 edges. For the nodes, we are left with 24 required CHUNKs and 2 recommended CHUNKs.

Next, we discuss the centrality analysis for OS3604 in Tables 4.8 and 4.9.

Table 4.8: Top 5 OS3604 CHUNKs based on betweenness centrality

CHUNK Name	Betweenness Centrality
Conducting Multiple Comparisons in ANOVA	31.15
Conducting One-Sample Hypothesis Tests	31.15
Conducting Paired-Sample Hypothesis Tests	31.15
Conducting Two-Sample Hypothesis Tests	31.15
Introduction to Confidence Intervals	31.15

Table 4.9: Top 5 OS3604 CHUNKs based on eigenvector centrality

CHUNK Name	Eigenvector Centrality
Conducting Multiple Comparisons in ANOVA	0.83
Conducting One-Sample Hypothesis Tests	0.83
Conducting Paired-Sample Hypothesis Tests	0.83
Conducting Two-Sample Hypothesis Tests	0.83
Introduction to Confidence Intervals	0.83

Through inspection, we clearly see there is no difference for the top five CHUNKs according to betweenness centrality and eigenvector centrality. We continue to notice a strong correlation between the CHUNKs in the network's core and the highest ranked CHUNKs in terms of centrality as each of these CHUNKs were contained within the network's core. Therefore, Prof. Isenhour would be able to effectively design future courses emphasizing the use of these required CHUNKs.

In the next chapter, we summarize our findings from the statistical analysis approach and the network analysis approach. Further, we discuss recommendations for future work and the conclusion of the thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5: Future Work and Recommendations

The main objective of this research was to determine a user-centric approach to AESs by investigating user performance within the CHUNK Learning platform. We were primarily concerned with analyzing real CHUNK Learning user data from the live system's reports, specifically related to content completion and content views.

Our purpose was to assign each user with a competence level that improves the overall situation awareness of the AES and hence, the effectiveness of the teaching-learning process. The development of the STAR algorithm addresses many questions that were previously unanswered in the current model of CHUNK Learning especially concerning how users navigate through their coursework in the system. Moreover, the network science algorithm to analyze connections between users and CHUNKs grouped and ranked educational content in terms of importance, providing another useful system tool to both instructors and students.

While both algorithms can be viewed as successful, each one still faces some major challenges that prevent the output of precise results that enable perfectly tailored recommendations to improve student learning. In this chapter, we present a summary of the main takeaways from our analysis in Section 5.1, discuss future enhancements to both algorithms in Section 5.2 and then finalize with conclusions in Section 5.3.

5.1 Summary

The statistical analysis of the CHUNK Learning data revealed some noticeable trends specific to each course. For MA4027, we found that there was better performance in VC scores compared to CR scores, which clearly correlates to the course's flexible position towards CHUNK completion. In OS3307, we observed the profound effect of a single CHUNK user in the VH/VH group. User 45's high VC score had a highly negative effect towards the other students' VC scores as 56% of the class achieved a low rating or worse. OS3604 achieved the best results out of the three courses as 58% of the class had a STAR level of 3 or better. This result supports the trend that courses which enforced CHUNK completion performed better than those that did not. In general, the STAR Rating System easily segregated eight students who earned a STAR level 1 out of the 68 students in the three courses.

The network science analysis showed mostly consistent results among the three courses. For MA4027, three recommended CHUNKs were found in the network's core. This was surprising because we initially believed the core to feature CHUNKs from the required CHUNK list. In contrast, the centrality analysis revealed only required CHUNKs but there was overlap between the two methods. The results for OS3307 and OS3604 were closely related as both courses exclusively featured required CHUNKs when ranking CHUNKs according to k-core and centrality. Thus, the network analysis results were indeed relevant and could be used to determine which CHUNKs were most important in each respective course.

5.2 Future Work

In this section, we present a number of suggestions to improve the methodology in the thesis. We propose new ways to enhance the STAR algorithm In Section 5.2.1 and then discuss improvements for the network analysis approach in Section 5.2.2.

5.2.1 STAR Algorithm Improvements

First, the algorithm may benefit by introducing a reference score for CR and VC. In terms of CR, we can use the instructor's number of required CHUNKs. In terms of VC, it can be the instructor's guidance for number of views per specific CHUNK. For example, it may be reasonable to suggest two views per activity of a required CHUNKlet within a CHUNK.

Next, we can weigh CR and VC points to potentially give one score more value than the other. Currently, both scores have an equal weight, but an instructor might value CR more than VC. For example, an instructor may choose to apply the following weights: $w_{CR} = 0.67$ and $w_{VC} = 0.33$. By increasing the weight of CR over VC, students will be given a STAR score that is more dependent upon how many CHUNKs they complete in the course.

Further, we are interested in how to best incorporate different cognitive parameters into an AES. Our methodology introduced two groups, VH/VL and VL/VH, that were deemed the highest risk for the Category 1 and Category 2 errors identified in Section 3.1.3. A more in-depth analysis of this research may uncover that these groups are related to working memory capacity. Specifically, we would like to prove if the VH/VL group is connected to a high working memory capacity and if the VL/VH group is connected a to a low working memory capacity.

Moreover, we may determine new trends by analyzing the data at the CHUNKlet level within each course. We can compare STAR groups and levels to those achieved at the CHUNK level to discover if there are any discrepancies between the two.

5.2.2 Network Science Algorithm Improvements

We can possibly enhance the network science algorithm and its results through the following changes. First, we can incorporate additional network science parameters to find the optimal selection of CHUNKs within a course. For instance, we can calculate each course network's modularity and then run community detection to separate CHUNKs into multiple communities. This would possibly correlate to the results of the k-core analysis and substantiate which CHUNK group is dominant in the network. Some other possible parameters to consider for individual CHUNK importance are: closeness centrality and eccentricity.

Next, we can also include additional edges within the CHUNK-User network to form a more robust network for analysis. One approach is to create edges between nodes based on course prerequisites, which can be further tailored by instructors.

Lastly, we can explore the use of network analysis computer programs other than Gephi. One of these programs, Cytoscape, is an open-source platform with various applications specifically designed to visualize and analyze networks just like the CHUNK-User networks we developed [21]. New programs may provide deeper insight into modeling student actions within CHUNK Learning.

5.3 Conclusions

The results of this project enabled CHUNK Learning to identify each user by a STAR group and level which represented some form of a competence level. While the analysis was useful for a homogeneous group of graduate-level students, we must be able to apply these algorithms to all levels of educations including secondary education and undergraduate education. We can then gain further understanding of different patterns of learning at varying ages. It is important to analyze students at the undergraduate level, especially those in their first and second years as these students may lack the discipline and motivation to learn from an AES.

Moreover, we can prevent errors from affecting our CHUNK Learning user data by installing certain restrictions in the system. These restrictions may be able to reduce human errors such as the Category 1 and Category 2 errors referenced in this thesis from impacting the STAR algorithm and the network science algorithm. For instance, we can prevent users from clicking complete on CHUNKs without viewing activities by forcing them to spend the recommended amount of time in each CHUNK.

Lastly, CR/VC scores were highly dependent upon instructors ensuring that students knew how to navigate and explore course content in the CHUNK Learning Explorer. Future studies must ensure that each user has attended a class or at least completed a tutorial on CHUNK Learning to prevent future errors from diminishing the quality of the data.

APPENDIX: Supplementary CHUNK Data

A.1 Required CHUNK list

In this section, we list the required CHUNKs for each course. These tables represent a plethora of learning content that the instructors presented to their students, varying on a week-by-week basis.

Table A.1: MA4027 required CHUNK list

Week	**CHUNK Description**
1	Gephi
	Overleaf for Documents Containing Mathematical Expressions
	MA 4027 Syllabus, Homework & Schedule
	Narrated Assessments
	Network Profile Summary (NPS)
2	Standard Terminology in Graph Theory
	Common Classes of Graphs
	Distance in Graphs
	Neighborhoods in Graphs
3	The Degree of a Vertex and Counting Friendships
	Regular Graphs
	Degree Sequences
4	The Definition of Isomorphism
	Isomorphism as a Relation
	Isomorphism Proofs
5	Bridges
	Trees
6	Cut-Vertices
	Blocks
7	Eulerian Graphs
	Hamiltonian Graphs
8	Vertex Colorings
9	Planar Graphs
	Embedding Graphs on Surfaces

Table A.2: OS3307 required CHUNK list

Week	**CHUNK Description**
1	Introduction to Modeling Practices for Computing
	Selecting a Statistical Programming Environment
	Introduction to CoCalc.com
	Getting Started: The Python Programming Environment
	Introduction to Statistical Modeling and Simulation with Python
2	Using Tabular Summaries to Describe Data
	Using Visual Displays to Describe Data
	Using Numerical Summaries to Describe Data
3	Introduction to Probability
	Understanding Conditional Probability and Bayes' Theorem
	Understanding Discrete Data and Distributions
	Understanding Continuous Data and Distributions
4	Modeling with Discrete Distributions
	Modeling with Continuous Distributions
5	Constructing a Point Estimate
	Introduction to Confidence Intervals
	Estimating the Mean of a Population
	Estimating the Proportion of a Population
	Introduction to Hypothesis Testing
	Conducting One-Sample Hypothesis Tests
6	Introduction to Simple Linear Regression
	Modeling Using Simple Linear Regression
	Using Linear Regression to Make Predictions
	Assessing Adequacy of a Linear Regression Model
	Introduction to Multiple Linear Regression
	Constructing Regression Models with Categorical Variables
7	Introduction to Modeling and Simulation
	Monte Carlo Simulation
	Reliability Modeling
8	Introduction to Discrete Event Simulation
	Transforming Event Graphs into Discrete Event Simulations
	Generating Statistical Models from a Discrete Event Simulation
9	Introduction to Queuing Models
	Computing Measures of Performance in Queuing Models
	Measuring Network Performance
10	Introduction to Design of Simulation Experiments
	Creating Full Factorial Designs
	Creating Fractional Factorial Designs

Table A.3: OS3604 required CHUNK list

Week	CHUNK Description
1	Introduction to Statistics and Data Analysis
	Selecting a Statistical Programming Environment
	Using Visual Displays to Describe Data
	Using Numerical Summaries to Describe Data
2	Point Estimation
	Introduction to Confidence Intervals
	Estimating the Mean of a Population
	Estimating the Proportion of a Population
	Estimating Population Variance
3	Introduction to Hypothesis Testing
	Conducting One-Sample Hypothesis Tests
4	Conducting Two-Sample Hypothesis Tests
	Conducting Paired-Sample Hypothesis Tests
5	Conducting Single-Factor Analysis of Variance
	Conducting Multiple Comparisons in ANOVA
6	Conducting Two-Factor Analysis of Variance
	Conducting Multi-Factor Analysis of Variance
7	Introduction to Simple Linear Regression
	Modeling using Least Squares Regression
	Using Linear Regression to Make Predictions
8	Assessing Adequacy of a Regression Model
	Introduction to Multiple Linear Regression
	Introduction to Binomial Logistic Regression
9	Conducting Analysis Using Categorical Data
	Constructing Regression Models with Categorical Variables

THIS PAGE INTENTIONALLY LEFT BLANK

List of References

[1] R. Gera, M. Isenhour, D. Bartolf, and S. Tick, "Chunk: Curated heuristic using a network of knowledge," in *The Fifth International Conference on Human and Social Analytics. HUSO*, 2019.

[2] D. J. Camacho and J. M. Legare, "Shifting gears in the classroom—movement toward personalized learning and competency-based education," *The Journal of Competency-Based Education*, vol. 1, no. 4, pp. 151–156, 2016.

[3] R. C. Clark and R. E. Mayer, *E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. Hoboken, NJ: John Wiley & Sons, 2016.

[4] A. Siddique, Q. S. Durrani, and H. A. Naqvi, "Developing adaptive e-learning environment using cognitive and noncognitive parameters," *Journal of Educational Computing Research*, vol. 57, no. 4, pp. 811–845, 2019.

[5] "Adaptive learning," Feb 2020. [Online]. Available: https://www.dreambox.com/adaptive-learning

[6] S.-L. Huang and J.-H. Shiu, "A user-centric adaptive learning system for e-learning 2.0." *Educational Technology & Society*, vol. 15, no. 3, pp. 214–225, 2012.

[7] N.-N. Manochehr *et al.*, "The influence of learning styles on learners in e-learning environments: An empirical study," *Computers in Higher Education Economics Review*, vol. 18, no. 1, pp. 10–14, 2006.

[8] A. Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.

[9] L. Ma, L. Chang, X. Chen, and R. Zhou, "Working memory test battery for young adults: Computerized working memory assessment," *PLOS One*, vol. 12, no. 3, 2017.

[10] N. Tsianos, P. Germanakos, Z. Lekkas, C. Mourlas, and G. Samaras, "Working memory span and e-learning: The effect of personalization techniques on learners' performance," in *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 2010, pp. 64–74.

[11] A. Gardner, "The viability of online competency based education: An organizational analysis of the impending paradigm shift," *The Journal of Competency-Based Education*, vol. 2, no. 4, p. e01055, 2017.

[12] W. G. Spady, "Competency based education: A bandwagon in search of a definition," *Educational researcher*, vol. 6, no. 1, pp. 9–14, 1977.

[13] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Boston, MA: Cengage Learning, 2011.

[14] G. Chartrand and P. Zhang, *A First Course in Graph Theory*. North Chelmsford, MA: Courier Corporation, 2013.

[15] M. Newman, *Networks: An introduction*. Oxford, England: Oxford University Press, 2010.

[16] S. P. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.

[17] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi, "Structure and dynamics of core/periphery networks," *Journal of Complex Networks*, vol. 1, no. 2, pp. 93–123, 2013.

[18] J.-D. Cleven, "A multilayer network approach for real-time adaptive personalized learning," Naval Postgraduate School, Monterey, CA, United States, Tech. Rep., 2018.

[19] M. Isenhour, "For content authors: Chunklets." [Online]. Available: https://wiki.nps.edu/display/CHUNKL/ForContentAuthors:CHUNKlets

[20] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Third international AAAI conference on weblogs and social media*, 2009.

[21] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome research*, vol. 13, no. 11, pp. 2498–2504, 2003.

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California