2021-06

# A RECOMMENDER MODEL USING SOCIAL TIE STRENGTH FOR THE CHUNK LEARNING SYSTEM

Critchley, Matthew F.

Monterey, CA; Naval Postgraduate School

http://hdl.handle.net/10945/67690

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

# THESIS

**A RECOMMENDER MODEL USING SOCIAL TIE STRENGTH FOR THE CHUNK LEARNING SYSTEM**

by

Matthew F. Critchley

June 2021

| | |
|---|---|
| Thesis Advisor: | Ralucca Gera |
| Second Reader: | D'Marie E. Bartolf |

**Approved for public release. Distribution is unlimited.**

THIS PAGE INTENTIONALLY LEFT BLANK

| REPORT DOCUMENTATION PAGE | *Form Approved OMB No. 0704-0188* |
|---|---|



| **1. AGENCY USE ONLY** ***(Leave blank)*** | **2. REPORT DATE** June 2021 | **3. REPORT TYPE AND DATES COVERED** Master's thesis |
|---|---|---|

| | |
|---|---|
| **4. TITLE AND SUBTITLE** A RECOMMENDER MODEL USING SOCIAL TIE STRENGTH FOR THE CHUNK LEARNING SYSTEM | **5. FUNDING NUMBERS** |
| **6. AUTHOR(S)** Matthew F. Critchley | |
| **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(E**S) 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)**

With the onset of COVID-19, rising tuition costs, and technological advancements, online courses have become a pervasive medium through which education is conducted. Currently, several online educational services tailor education to students through various methods of recommender models. One such system, the Curated Heuristic Using a Network of Knowledge (CHUNK) Learning, developed at the Naval Postgraduate School, uses a recommender system that relies on user profile attributes. We propose a complementary recommendation system to expand upon CHUNK's current recommender method by incorporating implicit recommendations from a user's social network based on tie strength between learners. In this work, we create a synthetic social network of learners and calculate the Jaccard Index and Pearson Correlation Coefficient similarity values to distinguish between strong and weak social ties. These tie classifications are then used to personalize content recommendations and expose users to greater breadth or depth of applicable knowledge based on current interests or job goals. We simulate recommendations for a user under different circumstances and show that our recommender system promotes the algorithmic formation of communities of learners on similar educational tracks. This promotes the social-emotional support for online learners that they may not currently receive and improves socialization within distance learning.

| | |
|---|---|
| **14. SUBJECT TERMS** network science, distance learning, CHUNK Learning, tie strength, recommender system, synthetic network | **15. NUMBER OF PAGES** 77 |
| | **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.**

**A RECOMMENDER MODEL USING SOCIAL TIE STRENGTH FOR THE CHUNK LEARNING SYSTEM**

Matthew F. Critchley
Ensign, United States Navy
BS, United States Naval Academy, 2020

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN APPLIED MATHEMATICS**

from the

**NAVAL POSTGRADUATE SCHOOL**
**June 2021**

Approved by: Ralucca Gera
Advisor

D'Marie E. Bartolf
Second Reader

Wei Kang
Chair, Department of Applied Mathematics

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

With the onset of COVID-19, rising tuition costs, and technological advancements, online courses have become a pervasive medium through which education is conducted. Currently, several online educational services tailor education to students through various methods of recommender models. One such system, the Curated Heuristic Using a Network of Knowledge (CHUNK) Learning, developed at the Naval Postgraduate School, uses a recommender system that relies on user profile attributes. We propose a complementary recommendation system to expand upon CHUNK's current recommender method by incorporating implicit recommendations from a user's social network based on tie strength between learners. In this work, we create a synthetic social network of learners and calculate the Jaccard Index and Pearson Correlation Coefficient similarity values to distinguish between strong and weak social ties. These tie classifications are then used to personalize content recommendations and expose users to greater breadth or depth of applicable knowledge based on current interests or job goals. We simulate recommendations for a user under different circumstances and show that our recommender system promotes the algorithmic formation of communities of learners on similar educational tracks. This promotes the social-emotional support for online learners that they may not currently receive and improves socialization within distance learning.

THIS PAGE INTENTIONALLY LEFT BLANK

# Table of Contents

# List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

# List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

# List of Acronyms and Abbreviations

| | |
|---|---|
| **AD** | Attacker-Defender |
| **DAD** | Defender-Attacker-Defender |
| **DOD** | Department of Defense |
| **GN** | Girvan-Newman community detection algorithm |
| **NPS** | Naval Postgraduate School |
| **USN** | U.S. Navy |
| **USG** | United States government |
| **UoI** | User of Interest |

THIS PAGE INTENTIONALLY LEFT BLANK

# Executive Summary

I had the familiar experience of pursuing an education during the lockdowns induced by COVID-19. Throughout my studies, conducted through distance learning platforms, I recognized a lack of socialization with my peers, and the difficulty of collaborating with others. Moreover, it proved exceedingly difficult to learn from the advice and experiences of my peers to assist in my academic and professional career.

The CHUNK Learning system, developed at the Naval Postgraduate School, is a distance learning platform that arranges educational content in a non-linear structure and provides a platform for personalized online instruction. In this work, we develop a recommender system to augment the CHUNK learning platform in an effort to connect learners as they progress in the online learning environment. Our proposed recommender system relies entirely on a user's social network to generate content recommendations. Through this construction we believe we can increase socialization in a distance learning environment and improve content recommendations by using the expertise of peers to drive an individual's learning.

Our system leverages network structure and creates three networks to fuel recommendations. The first is a synthetic social network that links individuals in a distance learning environment together. We use the Barabasi-Albert "scale free" model and a Stochastic Block Model to replicate the core-periphery structure observed in Massive Open Online Courses (MOOCs). Our second network maps the content from the CHUNK learning database and creates a content network where individual lessons called chunklets, and chapters called chunks, are connected based on their structure within the database. Our final network ties these previous two networks together by connecting users to the content they have completed.

Using the Jaccard Index and Pearson Correlation Coefficient we calculate node similarity between users in our synthetic social network. Based on these values we then classify pairs of users as either "strong" or "weak" ties. From these classifications, we can recommend material from strong ties to develop a user's depth of knowledge or we can recommend material from weak ties to develop a user's breadth of knowledge.

We show that through the recommended content sub-networks our recommendations are more precise, accurate, and personally tailored than a recommender system based on dynamic keyword attachment. We also show that from recommendations we can algorithmically promote the formation of communities within the learning environment that consist of users on similar educational tracks, interested in similar material, or pursuing a career in a specific area. We conclude with a discussion of possible extensions of this work.

# Acknowledgments

I would like to thank my advisor Dr. Ralucca Gera and second reader Mrs. D'Marie Bartolf for their assistance throughout this thesis. They both volunteered several hours throughout the past six months to help me formulate my topic and execute the methodology. Additionally, their help in the preparation of this paper and my presentation was immense and I thank them for teaching me so much throughout the process.

I would also like to extend my thanks to the Naval Postgraduate School and the Applied Mathematics Department for allowing me to earn a master's degree and continue my education. My peers in the Applied Mathematics Department were also of great help throughout my time at NPS and I would like to specifically thank Maj Andrew Dietz and ENS Tommy Reeder for their assistance and advice throughout the thesis process as well.

Finally, I would like to thank my family back home for their support and love, as well as my friends, colleagues, and fellow service members at NPS, the Monterey area, and around the world, for making this unique year academically rewarding.

THIS PAGE INTENTIONALLY LEFT BLANK

# CHAPTER 1: Introduction

More than one billion children are at risk of falling behind in school due to restrictions surrounding COVID-19(UNICEF 2020). All levels of academia have been affected by mandated shutdowns and quarantines, including higher education. The National Student Clearing House Research Center reported that fall college enrollments dropped by 6.8%, 4.5 times larger than the pre-pandemic rate of decline (Causey et al. 2021). In addition to the COVID-19 pandemic, individuals are finding it increasingly difficult to attend college due to rising tuition costs.

Largely unaffected by change caused by the digital revolution, academia is beginning to see that a new future is quickly approaching. Even before the COVID-19 pandemic roiled academia “63.3% of higher education academic leaders agreed that online learning will become essential to their long-term strategy” (Lai et al. 2019). We expect that percentage to be much higher following COVID-19.

Current online learning platforms provide a viable way to enhance learning but suffer from pitfalls that result in adoption hesitancy. In this paper we propose a method to address one such pitfall, the lack of socialization. By developing a recommender model based on a user’s social network, we improve user exposure to relevant academic content, and connect users through the material they complete to promote socialization in the currently isolated distance learning experience.

## 1.1 Classroom of the Future

The classroom of the future will not be a single room with desks aligned in neatly oriented rows and columns, with a teacher standing in front of students occupying each desk. We envision the classroom of the future to be a network of students and teachers, studying from disparate locations, at a personalized pace, studying different topics at different times, all while fostering a social environment that leverages the advantages of social interaction with the efficiency of online learning.

Academia is one of the last areas to undergo a digital revolution, and with the emergence of COVID-19 induced lockdowns, academia's delayed digital revolution has arrived at an inflection point. Learning will take a new form, hybridizing instructor guided lessons with personalized educational journeys determined by a learner's interests and individual ownership.

Existing distance learning platforms allow users to engage with content and accomplish various goals like professional development, continuing education, certification, and attainment of a degree. These options allow students to gain access to various degrees of personally tailored educational tracks. However, the majority of existing programs fail to maximize inherent advantages of creating a non-linear, networked learning environment. This is one of the strengths of an emerging application in the field, the Curated Heuristic Using a Network of Knowledge (CHUNK) Learning (Gera et al. 2020).

Distance Learning (DL) through CHUNK provides an opportunity to empower learners and decentralize education in a way that enhances educational accessibility and quality. Despite the bright future ahead of education, there still exist issues in DL systems that give students and educators pause.

## 1.2 Social Interaction: A Distance Learning Shortcoming

One such weakness is the solitary nature of current DL platforms. Although online learning provides an efficient way to gain access to material, it hinders learners from developing a professional social network. How then can we combine the advantages DL has to efficiently personalize course interaction with the potential it has for student to student engagement?

In this work we use a user's social network to fuel a recommender model to promote social interaction and provide users with additional skills necessary in the 21st century. We believe the combination of our system with the underlying structure and personalized learning methods employed by the CHUNK Learning network will enhance user interest in distance learning, promote networks of life-long learners, and improve academia as a whole.

## 1.3 Overview of the Thesis

In Chapter 1 we introduce the issue of socialization within distance learning that we seek to address and improve. Then in Chapter 2, we define terms from network science and graph theory used in our recommender model and look at current research and theories to improve online learning. In Chapter 3 we outline the methodology in our recommender model and simulation. Then in Chapter 4, we present our results. In Chapter 5 we draw conclusions from the results and identify future areas of research.

THIS PAGE INTENTIONALLY LEFT BLANK

# CHAPTER 2:
# Background

In this chapter we introduce the networked learning system called "CHUNK Learning," educational theories underpinning networked learning, and establish common terms and definitions on network science- the mathematical basis of our analysis. Throughout the chapter we will introduce the findings of existing research works.

## 2.1 CHUNK Learning

The Curated Heuristic Using a Network of Knowledge (CHUNK) Learning is an online learning environment based on a theory of networked learning. Networked learning is defined by Caroline Haythornthwaite, as a method of instruction with the "pedagogical aim to enhance the use of network connectivity in support of educational outcomes" (Haythornthwaite 2019). CHUNK is "A personalized, adaptive, learning platform that breaks away from the predictable pattern of traditional education models and provides content delivery that respects the different capabilities, learning styles, and approaches to problem-solving of every learner" (Gera et al. 2020). CHUNK Learning breaks material into various modules called chunks. These chunks are arranged in a network structure to guide students through a course and can be thought of as chapters/ sections of a course syllabus. Each chunk is connected via edges to prompt students toward related material. These edges can be directed in order to provide students with a formal order of instruction, or undirected to simply suggest relevance to different topics. Figure 2.1 displays the home page layout, and several modules of the CHUNK Learning platform.

Within each chunk are different chunklets organized into four different categories: "Why", "How/What", "Methodology", and "Assessment." The chunklet is the building block of each Chunk and can consist of one or more activities. These activities can include any type of educational material such as videos, codes, lectures, etc. As a student progresses through a course they complete a required number of activities to finish a chunklet and a required number of chunklets to finish a chunk. For purposes of clarification we use the term CHUNK to refer to the entire learning platform, chunk to refer to chapter/sections, and chunklets to

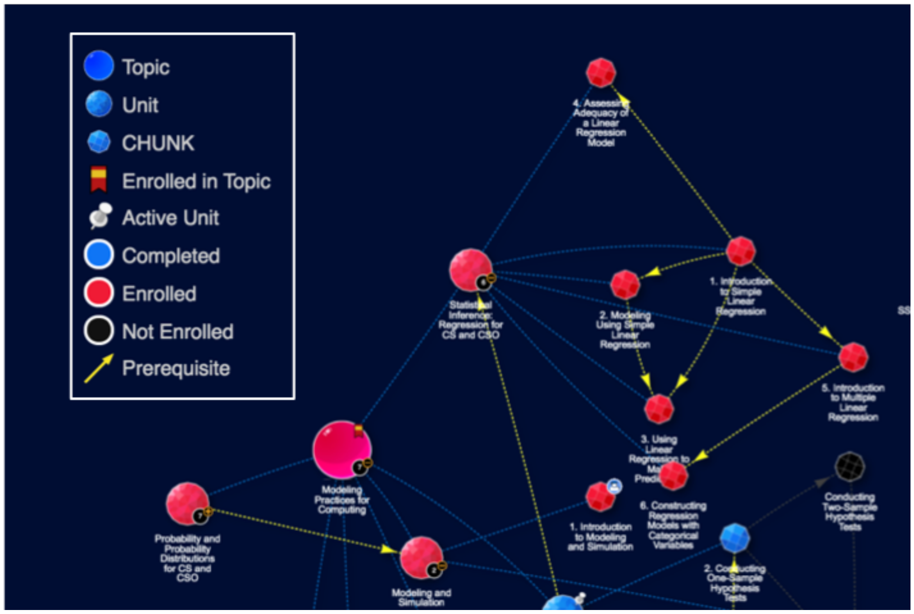


Figure 2.1. CHUNK structure

refer to the building block “Why” “How/What” “Methodology” and “Assessment” materials.

Besides the visual organization allowed by CHUNK Learning, the network structure also gives opportunity to utilize network science to implement recommender systems to guide users to new content. Currently CHUNK employs three recommender systems at the network (chunk level), chunklet level, and the activity level. For the purposes of this paper we are interested in the chunk and chunklet level recommender systems.

## 2.2 Connectivism, Community of Practice, Social Capital Theory

In this section we look at the learning theories underpinning networked learning and form the lens through which we seek to improve our network.

### 2.2.1 Connectivism

We believe society has largely relied on three different learning theories in our current educational system: behaviorism, cognitivism, and constructivism. Behaviorism theorizes that learning is about behavior change and that over time due to external factors, a response is learned. Cognitivism relates learning to computer memory, where learning is initially located in short term memory and overtime is coded into long-term recall. Constructivism takes the view that learners create knowledge from their own experiences. Despite the success of such theories on education structure in the past, recent changes caused by the digital revolution have altered the requirements of today's learning environment.

Connectivism, a learning theory introduced by George Siemens in 2004, focuses on the shifts caused by the digital revolution. Connectivisim focuses on several principles:

- Leaning and knowledge rests in diversity of opinions
- Learning is a process of connecting specialized nodes or information sources
- Learning may reside in non-human appliances
- Capacity to know more is more critical than what is currently known
- Nurturing and maintaining connections is needed to facilitate continual learning
- Ability to see connections between fields, ideas, and concepts is a core skill
- Currency (accurate, up-to-date knowledge) is the intent of all connectivist learning activities
- Decision making is itself a learning process. Choosing what to learn and the meaning of incoming information is seen through the lens of a shifting reality. While there is a right answer now, it may be wrong tomorrow due to alterations in the information climate affecting the decision. (Siemens 2005)

The principles of connectivism are a direct response to requirements imposed on individuals by the digital age. The digital revolution has increased the rate of information in our world, and this growth has continued at an exponential rate. The drastic increase in new information has caused half-life of knowledge to drop dramatically. The half-life of knowledge is the time frame in which knowledge is obtained until that information becomes obsolete. By one estimate, "the amount of knowledge in the world has doubled in the past 10 years and is doubling every 18 months" (Siemens 2005). Connectivism attempts to change the focus of education in order to deal with rapidly changing information and knowledge. The implications of this theory and the changes in educational structure are numerous and consequential. Since we cannot possibly learn everything, we must rely on a network of knowledge from our peers. Siemens points out "the capacity to form connections between sources of information, and thereby create useful information patterns, is required to learn in our knowledge economy" (Siemens 2005). Employing networked learning thus addresses the issues laid out by a digital society and allows learners to complement their learning with the skills noted by Stephenson and Siemens.

Similar to the second listed priority of connectivism above, is the idea of absorptive capacity, which is defined as "the ability of a firm to recognize the value of new, external information, assimilate it, and apply it" (Cohen and Levinthal 1990). We include this idea as another benefit networked learning incorporates in the education of individual users but also as a means for organizations to employ methods for improved organizational structure. Cohen and Levinthal state "the observation that the ideal knowledge structure for an organizational sub-unit should reflect only partially overlapping knowledge complemented by non-overlapping diverse knowledge" (Cohen and Levinthal 1990). This is exactly the structure of learning institutions, with different departments collaborating at times, but also furthering their knowledge of their area of expertise. We believe that in an online learning environment, we can leverage this structure to improve learning outcomes for students.

### 2.2.2 Community of Practice

The community of practice (CoP) is a widely cited theoretical frameworks in networked learning. In 2011, Wenger defined a CoP as a "learning partnership among people who find it useful to learn from and with each other about a particular domain. They use each other's experience of practice as a learning resource" (Smith et al. 2017). However, we find that a slightly different classification laid out by Henri and Pudelko may give more insight into communities in our online learning environment. There are four different classifications laid out by Henri and Pudelko, "communities of interest; goal oriented communities; a learner's community; and communities of practice" (Smith et al. 2017). The differentiation between these four classifications is based on the strength of ties between users in the community. Communities of interest possess the lowest cohesion as learning is strictly driven on an individual basis. Goal oriented communities have higher cohesion as they are driven by an external force to complete material in a predetermined time frame. A learner's community relies on a single authority, in most cases an instructor, for guidance in the learning process. Communities of practice have the highest level of cohesion as they consist of professionals who use their strong social ties to enhance their knowledge in order to complete similar tasks. In our paper we use this framework as a goal to achieve in connecting users within our network.

The types of communities outlined by the community of practice framework, rely on a rating of the "strength" of a tie between users of a community. Much discussion has occurred and continues to occur over what is classified as a learning tie (Haythornthwaite and De Laat 2010). In our network we do not give a definitive solution to determining a learning tie however, we do generalize the strength and weakness of ties between users based on their overlapping neighborhoods.

### 2.2.3 Social Capital Theory

The theory of social capital refers to "the inherent qualities and subsequent benefits that exist in or between social groups or networks due to the interaction among its members" (Venter 2019). Similar to connectivism, social capital theory states the belief in the inherent value of an individual's social network. Social capital theory creates a framework in online learning that distinguishes between the types of ties between students, and how such differences can lead to distinct benefits. Social capital theory "shows how different types of social ties in [personal learning environments] provide bonding and bridging social capital; the combination of which serves the learning project by providing for both strong ties in supportive relationships between students and weak ties with knowledge generation capabilities between previously unacquainted students" (Venter 2019). Learning environments encompass all the learning activities students engage in, whether they be formal or informal. In Venter's qualitative study he concludes that strong ties promote bonding social capital that causes students to engage in "learning together in reciprocal and supportive relationships while weak ties promote bridging capital that allows for the formation of bridges between strangers and the creation of an extended learning community" (Venter 2019). We utilize these two types of capital to highlight benefits learners receive from incorporating our social recommendation system.

Together the theories of connectivism, community of practice, and social capital theory act as the theoretical foundation for our improved distance learning social recommender system. We now shift to the mathematical foundation of our recommender system and introduce select network science and graph theory concepts needed for analysis.

## 2.3 Network Science and Graph Theory

This subsection covers the mathematical foundation of graph theory and network science we use in the analysis of our networks. We begin with the basic definition of a graph.

**Definition 2.3.1** *Graph*

> *an ordered pair of finite disjoint sets $(V, E)$ such that $E$ is a subset of the set $V^2$ of unordered pairs of $V$. The set $V$ is the set of vertices and $E$ is the set of edges. If $G$ is a graph, then $V = V(G)$ is the vertex set of $G$, and $E = E(G)$ is the edge set. An edge {x, y} is said to join the vertices x and y and is denoted by xy. Thus xy and yx mean exactly the same edge; the vertices x and y are the end vertices of this edge. (Bollobás 1998)*

In some graphs there are nodes and edges whose inclusion in a certain type of analysis is unnecessary. In such case, our attention is restricted to a portion of the graph, and so we define a subgraph as:

**Definition 2.3.2** *Subgraph*

> *a graph $H$ such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ and the assignment of endpoints to edges in $H$ is the same as in $G$. We then write $H \subseteq G$. (West et al. 2001)*

We point out here that throughout the paper, graph and network are terms used interchangeably, although generally, a network implies a higher level of complexity in the information associated with nodes and edges. Newman defines a network specifically as:

**Definition 2.3.3** *Network*

> *a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else. (Newman 2018)*

Another useful way to characterize a network is through its adjacency matrix. The adjacency matrix of a network is defined as:

**Definition 2.3.4** Adjacency Matrix

*an NxN matrix **A** whose elements $a_{ij}$ indicate whether node i is linked to node j. The elements $a_{ij}$ are given by:*

$$a_{ij} = \begin{cases} 1 & \textit{node i has an edge connecting it to j} \\ 0 & \textit{otherwise,} \end{cases}$$

We can expand our definition of network and adjacency matrix to account for weighted networks as well. A weighted network is a network where a weight is associated with each interaction, describing typically a measure of the "intensity" of the interaction. For weighted networks we can replace the value of 1 at each $a_{ij}$ with the weight of the edge $w_{ij}$ (Bianconi 2018).

We also deal with situations where the type of interaction between nodes is not always the same. Under these situations we define multilayer networks.

**Definition 2.3.5** Multilayer Network

*formed by several interacting networks of different edge types. A given multilayer formed by distinct M layers is formed by a set of M networks describing the interactions within each layer (Bianconi 2018). Mathematically, a multilayer network **M** is:*

$$\boldsymbol{M} = (Y, \vec{G})$$

*with Y being the set of layers*

$$Y = \alpha| \in 1, 2, ..., M$$

*and $\vec{G}$ the ordered list of networks characterizing interactions within each layer $\alpha = 1, 2, ..., M$*

$$\vec{G} = (G_1, G_2, ..., G_\alpha, ..., G_{\boldsymbol{M}})$$

*each subnetwork $G_\alpha$ is defined as:*

$$G_\alpha = (V_\alpha, E_\alpha).$$

*We note that each layer of the multilayer network is formed by the same set of nodes thus $V = V_\alpha \ \forall \ \alpha$. (Bianconi 2018)*

### 2.3.1 Network Topology

Once we establish the network, we are interested in analyzing, we then use different structures and measures to draw conclusions. We begin with a basic structure we can create in all networks called a path:

**Definition 2.3.6** *Path*

*a sequence of nodes such that every consecutive pair of them is connected by a link in the network. (Aleta and Moreno 2019)*

Aleta and Moreno note, an interesting and particularly useful application of a path is the random walk. A random walk is a path where each consecutive node is chosen at random from the adjacent nodes of the current node (note that a random walk allows nodes and edges to be repeated) (Aleta and Moreno 2019).

Next, we define one of the most widely measured statistics on nodes- node centralities. There are various types of centralities that all measure the "influence" a node has on the network. Here we define the three most common: degree, betweenness, and eigenvector centrality.

**Definition 2.3.7** *Degree Centrality* *measures the involvement of a vertex in a network by the number of vertices connected to it. (Newman 2018)*

**Definition 2.3.8** *Eigenvector Centrality*

*a degree centrality score proportional to the sum of the degree centrality scores of its neighbors. (Newman 2018)*

**Definition 2.3.9** *Betweenness Centrality*

*the ability of a vertex to play a "broker" role in the network by measuring how well it lies on the shortest paths connecting other vertices. (Orman et al. 2013)*

One study of an online class by Russo and Koesten looked at how degree centrality was related to cognitive learning outcomes but not affective learning outcomes (Russo and Koesten 2005). In the study they defined cognitive learning outcomes as knowledge retention, while affective learning was defined as learning that influenced further investigation and motivation on the learned material. By making a distinction between in degree and out degree, Russo and Koesten defined two types of centralities that they called *centrality* (out-degree) and *prestige* (in- degree). Centrality is a measure of user's influence and represents a "major channel of relational information." Whereas prestige represents whether other nodes "seek out a particular actor in a social network". A similar finding was reported by Lai et al. stating "individuals who played the 'Center' role were more likely to earn a higher learning achievement in the class" (Lai et al. 2019).

Moving beyond centrality, a useful measure related to betweenness centrality is average path length. Newman defines the average path length as:

**Definition 2.3.10** *Average Path Length*

*the mean geodesic or shortest-path distance between pairs of vertices. (Newman 2018)*

This measure encompasses all nodes in a network and is a global measure of the general distance between two nodes. However, when there are certain nodes or groups of nodes that are disconnected from the larger network, this measure can be slightly misleading since these disconnected nodes cannot be reached. Another useful description of a network is to classify the number of connected components. If every node in a network is connected to every other node by some path then the network is said to have one component, otherwise it will have multiple components. Newman defines the component in an undirected network as:

**Definition 2.3.11** *Component*

> *a maximal subset of vertices such that each is reachable by some path from each of the others. (Newman 2018)*

In addition to average path length, the size of a network can also be quantified by the network diameter.

**Definition 2.3.12** *Diameter*

> *the length of the longest finite geodesic path anywhere in the network. (Newman 2018)*

Moving on from size of networks another useful measure for our work is how tightly connected a network is. Network density is defined by Newman as:

**Definition 2.3.13** *Density*

> *the fraction of edges that are actually present, out of the total number of possible edges. (Newman 2018)*

Based on this definition network density is a measure that ranges from 0 to 1, where a network density of 0 represents a network in which no nodes have any edges to each other, while a network density of 1 means every node is adjacent to every other node. We can use this measure of density on subgraphs where a subgraph that has a density of 1 is referred to as a "clique."

Similar to density, the clustering coefficient is another useful measure that allows analysis on the connectedness of each node's neighbors in a network.

**Definition 2.3.14** *Clustering Coefficient*

> *measures the average probability that two neighbors of a vertex are themselves neighbors:*
> $$C = \frac{(\textit{number of triangles}) \times 3}{\textit{number of connected triples}},$$
> *where connected triples means three vertices, $uvw$, with edges $(u, v)$ and $(v, w)$. (Newman 2018)*

Density, and clustering coefficients can lend valuable insight into *homophily*, defined as "a propensity for similar actors to be disproportionately connected in a relation of interest" (Grunspan et al. 2014). Homophily is the basis of many collaborative filtering recommender systems which we define in Section 2.5.

### 2.3.2 Community Detection

A major aspect of network analysis relies on algorithms that detect subsets of nodes, or communities, within the entire network. There is no widely accepted definition of a community among network scientists as many believe that such a definition would hamper the creativity and development of detection algorithms. In this paper we use the community detection based on the definition by Radicchi et al. as a foundation.

**Definition 2.3.15** *Communities*

> *a subset of vertices within the graph such that connections between vertices within the community are denser than connections with the rest of the network (Radicchi et al. 2004).*

*Further classification of communities can be either weak or strong. The subgraph, $V$, is a community in a weak sense if the sum of all of the degrees within $V$ is greater than the sum of the degrees towards vertices outside of $V$, where $k_i^{in}(V)$ and $k_i^{out}(V)$ represent the degrees of the vertices inside and outside of the community respectively. Symbolically this relationship is represented as:*

$$\sum_{i \in V} k_i^{in}(V) > \sum_{i \in V} k_i^{out}(V).$$

*Radicchi et al. consider a subgraph $V$ a community in a strong sense if each vertex has more connections, $k_i$, inside the community, $k_i^{in}(V)$, than the vertex has with the remainder of the network outside of the community, $k_i^{out}(V)$.*

$$k_i^{in}(V) > k_i^{out}(V), \forall i \in V.$$

An often used method in determining communities is centered on the modularity value of a network.

**Definition 2.3.16** *Network Modularity*

*the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random (Newman 2006).*

*For network $C$ consisting of $L$ edges and adjacency matrix $A$ mathematically this is:*

$$M_c = \frac{1}{2L} \sum_{(i,j) \in C_c} (A_{ij} - p_{ij})$$

*where $p_{ij}$ is the expected number of links between nodes $i$ and $j$ if the network were wired at random (Barabasi 2016).*

In our work we utilize the Louvain Algorithm to determine network modularity for our recommendation sub-networks. The Louvain Algorithm seeks to sort nodes into communities by achieving the maximum modularity value throughout the network. The algorithm does so by first assigning each node to an individual community, and then moving each node into a neighbor's community to maximize the modularity value. This process is repeated until no further improvement can be made. Then each community is mapped to a single node in a new network and the same process is applied. This continues until no further improvement can be made.

### 2.3.3 Node Similarity Metrics

Similar to community detection among an entire network, we also use a measure of similarity between pairwise combinations of nodes. In our work we use two different methods of determining node similarity: the Jaccard Index, and Pearson Correlation Coefficient.

**Definition 2.3.17** *Jaccard Index*

> *the quantity of neighbors two nodes share normalized by the size of their full neighborhood. The neighbors of node i is denoted as $N(i)$. Thus, two nodes are more similar if they possess edges to the same users. We calculate the Jaccard index $J_{u,v}$ using the union and intersection of neighbors of the two nodes:*

$$J_{i,j} = \frac{|N(i) \bigcap N(j)|}{|N(i) \bigcup N(j)|}. \tag{2.1}$$

We note that the Jaccard Index ranges from 0 to 1 with 0 representing a scenario in which two nodes share no neighbors and 1 representing a scenario where two nodes possess the edges to the exact same set of vertices.

**Definition 2.3.18** *Pearson Correlation Coefficient*

> *a measure that compares the number of common neighbors between two nodes to the expected number of common neighbors if the edges in the network were determined at random. We calculate the Pearson Correlation Coefficient (PCC), denoted $r_{i,j}$, as:*

$$r_{i,j} = \frac{\sum_k [a_{ik} - < a_i >][a_{jk} - < a_j >]}{\sqrt{\sum_k [a_{ik} - < a_i >]^2}\sqrt{\sum_k [a_{jk} - < a_j >]^2}} \tag{2.2}$$

In Equation 2.2, $a_{ik}$ refers to the element in the $i$-th row and $k$-th column of the network's adjacency matrix. This is the number of edges between node $i$ and node $k$. The term $< a_i >$ is the average value of the $i$-th row in the adjacency matrix and is $\frac{\deg(i)}{n}$ where n is the total nodes in the network. We sum over all $k$ nodes in the network to find the PCC between nodes $i$ and $j$. We note that the PCC ranges from -1 to 1. Negative values represent two nodes who have less neighbors than expected by chance, and positive values represent two nodes who have more neighbors than expected by chance. Thus pairs of nodes with positive PCC are considered more similar than pairs with negative PCC.

## 2.4 Network Models

In this section we describe several types of network models developed in the field of network science. We begin with the random network model called the Erdos-Renyi (ER) network (Barabasi 2016). The goal of any model is to accurately resemble real life networks. The ER model does so by assuming that connections between nodes are created at random. Erdos and Renyi initially used a model $G(N, L)$ where $N$ is the number of nodes in a network and $L$ is the number of edges but later, Edgar Gilbert introduced a random graph model G(N,p) where $p$ is the probability each pair of $N$ labeled nodes is connected. This model has proven more useful in determining network characteristics (Barabasi 2016). Observed in many real-world networks, the small worlds property, is the phenomenon that the distance between any two randomly chosen nodes in a network is short (Barabasi 2016). The small worlds property exists in ER networks however, they are missing another property observed in many real networks— high clustering.

The Watts-Strogatz (WS) model improves on the ER random graphs by including both the small worlds phenomenon and high clustering. The WS model begins with a ring of nodes connected to their immediate neighbors, then with probability $p$, each link is rewired randomly to another node in the network. The rewiring across the network drastically decreases the diameter of the network while keeping the clustering coefficient large. We note that as $p \Rightarrow 1$ the WS becomes the ER graph. However, for $0 < p < .1$ the WS model holds the small worlds and high clustering properties observed in real network (Barabasi 2016). The WS model however, does not successfully take into account that across many networks the degree of nodes follows a power law distribution of the general form $p_k = Ce^{-\lambda k}$.

The Barabasi-Albert (BA) model takes another step in replicating real-world network structure by incorporating a power law degree distribution for nodes in a network. The presence of this power law distribution, often referred to as "scale free," incorporates the presence of hub nodes. These nodes often have orders of magnitude higher degree than most other nodes in the network. They are present in several different real-world networks including social networks. The BA model does so by connecting nodes via preferential attachment, meaning the probability a link between a new node and an existing node depends on the existing node's degree (Barabasi 2016). We use this method as one way to create our synthetic social network.

We also make use of the stochastic block model (SBM) in order to produce a network with structured communities. The model takes in $n$ nodes, a partition of vertices $\{C_1, C_2, ..., C_r\}$, and a symmetric $r \times r$ matrix $P$ of edge probabilities. We use this model to distinguish between strong and weak ties of a core-periphery structure as observed in the Massive Open Online Courses (MOOCs) studied by Kellogg, Booth, and Oliver (Kellogg et al. 2014). In networks, various intermediate scale structures arise mostly concerning the formation of communities within a larger network. One such structure is the core-periphery structure which can be qualitatively described as "a densely connected core of nodes and sparsely connected collection of peripheral nodes [where] nodes in a core are also reasonably connected to those in a network's periphery" (Rombach et al. 2017). For both our BA and SBM synthetic networks we use the NetworkX Python library.

## 2.5 Recommender Models

Recommender systems are present in many of the applications we use every day. Experience with them most likely occurs when shopping online or viewing movie suggestions. Here we look at four of the most common types of recommender models: *Collaborative Filtering (CF)*, *Content-based recommending*, *Community-based recommending*, and *hybrid approaches*. Due to the wide range of applications of recommender models, in this section we refer to the material they recommend as items. These items may be entertainment, products, etc., or in our case educational material.

### 2.5.1 Collaborative Filtering

Collaborative filtering is considered by many as the most popular method employed in recommender systems, and can be broken into two methods. The first uses the preferences of similar users to recommend items while the second recommends items similar to those they have consumed in the past. The similarity between users or items is determined can be calculated by various methods such as cosine-based similarity (Lu et al. 2015). Issues may arise, however, since these methods only take into account users who have rated both items. Thus, some items with few ratings may incorrectly express high similarity with other items. To account for this, Jaccard metric weighting has been employed (Lu et al. 2015).

### 2.5.2 Content-Based Recommending

Content-based recommending removes the user similarity determination of collaborative filtering and only looks at the similarity between items the user has consumed or rated positively in the past. Items are characterized by their features and recommendation is based on the extent of shared features between the item and the preferences of a user from past items. Two techniques are widely used to generate recommendations, cosine similarity and statistical and machine learning using user history as the training data (Lu et al. 2015). This type of recommending presents users with items that share resemblance to other items they have consumed and/or rated in the past (Ricci et al. 2015). Currently, this method is employed in CHUNK Learning by pairing attributes attached to user profiles with attributes attached to chunks and chunklets.

### 2.5.3 Community-Based Recommending

This recommender system, often referred to as social recommendation, uses the preferences of a user's friends. This system gathers information about the social relations of a user and recommends items that their friends have rated highly and/or consumed (Ricci et al. 2015). The notion of trust is highly important in this type of recommender system as studies have shown "there is positive correlation between trust and user similarity in online communities" (Lu et al. 2015). An emerging field of research, trust has been substituted by other data measures present in open social networks such as social tags and co-authorship relations (Lu et al. 2015).

### 2.5.4 Hybrid Approaches to Recommending

The final category we define here is hybrid recommendation models. Hybrid approaches take a combination of the aforementioned techniques in order to minimize the disadvantages of each method. We consider our recommender model a hybrid approach to recommendation by combining the three methods.

## 2.6 Tie Strength in Recommender Models

Due to the widespread use of recommender models in society, there has been much research and discussion on the various types of recommender models discussed in Section 2.5. Research conducted on recommender models based on tie strength is a growing area of interest. One paper looking into the effectiveness of tie based recommender models was conducted by Wang et al. (Wang et al. 2016). In their work they investigate four real-world data sets and distinguishing between weak and strong ties based on a node's Jaccard coefficient and rank content based on whether content was recommended by strong or weak ties. The data sets investigated were a co-authorship citation network, a DVD rating and user trust community named Ciao, a Chinese forum social networking site Douban, and consumer reviews from Epinions. In each of the cases they developed two models which separately ranked strong or weak ties higher. Their findings suggest that both models improved performance but the model recommending items based on weak social ties performed considerably better than existing recommender models.

In this work we expand upon the findings of Wang et al. by developing and applying a similar recommender system to CHUNK Learning by utilizing a mixed ranking of strong and weak social ties to recommend educational content to users. From this system we plan to demonstrate that by distinguishing between strong and weak ties, recommendations become more efficient by exposing students to a greater depth of material in their field of interest as well as a wider breadth of material in topics tangentially related to their current interests. The combination of which, maximizes the advantages of the connectivism learning theory and prepares students for today's working environment.

THIS PAGE INTENTIONALLY LEFT BLANK

# CHAPTER 3: Methodology

In this chapter, we state the problem and discuss our solution technique. Finally, we provide our methodology for solving the problem.

## 3.1 Problem Statement

As outlined by social capital theory, we recognize the advantages of utilizing strong and weak ties in knowledge generation. To our knowledge there are no such applications to recommender models tailored for online learning environments. As discussed in Section 2.5, Wang et al. demonstrated various improvement measures of social recommendation with strong and weak ties compared to other recommender models. Our methodology provides an opportunity for a complementary recommender system to existing methods employed in distance learning environments.

## 3.2 Our Recommender System

In the existing CHUNK Learning system there are two types of users, directed and exploratory. A directed user is one who is currently enrolled in a course and is directed by an instructor to complete certain chunks for credit and an exploratory user is one who is learning on their own, driven by their own interests or needs to up-skill based on that user's existing skills. In this work, our recommender system focuses on improving the choice of content presented to an exploratory user of interest (UOI). We utilize three separate networks to determine recommendations as we introduce next.

### 3.2.1 Social Network

The first network, a user's social network is used to take advantage of a user's connections with others to drive content recommendation. Social capital theory is predicated upon the idea that strong social ties promote binding capital, while weak social ties promote bridging capital. In the context of distance learning environments, binding capital corresponds to the benefit gained from interaction between tightly connected individuals in the same discipline.

Bridging capital, on the other hand, connects users who are weakly connected but have the potential to expose individuals to new information and ways of thinking. In addition to improved recommendations, we believe utilizing these ties will act as a catalyst to promote socialization in distance learning environments.

We begin by creating a synthetic network of users resembling a real-life social network. For our work, we create two separate cases of a social network for a deeper analysis: for the first case a Barabasi-Albert (BA) Scale Free model, and for a second case a stochastic block model (SBM). We use the BA model to approximate the structure and degree distribution of real world social networks (Barabasi 2016), and the SBM to replicate core-periphery structure observed in Massive Open Online Courses (MOOCs) (Kellogg et al. 2014).

**Case 1: Barabasi Albert model**

The user social network, $G_u$, consists of the vertex set $V_u$ and the edge set $E_u$. The vertex set contains a node $v_i$ for each user $i$. The users in this network can consist of both exploratory and directed users. Edges between users are determined by the NetworkX built-in barabasi_albert_graph function. This function returns a random graph with a built in preferential attachment method outlined by the Barabasi Albert model (Albert and Barabási 2002). The function takes in two values $n$ and $m$, where $n$ is the number of nodes in the final network, and $m$ is the number of edges to attach from a new node to existing nodes. For our experiment, we set $n = 500$ students, starting with an initial set of $m = 10$ edges.

Next we add a new user to the network of users, under two separate scenarios for the BA model. This added node represents a single User of Interest (UOI) who is used in our simulations of chunklet completion for the following scenarios:

- the UOI node is connected to 1 existing user, and
- the UOI node is connected to 10 existing users

When our UOI is connected to 1 existing user, we replicate an individual who is new to a distance learning environment, an issue for recommender systems referred to as the "cold-start problem." When our UOI is connected to 10 existing users, we look at the effect of recommendations on a user who is well established in the distance learning environment.

**Case 2: Stochastic Block model**

Our stochastic block model is another built-in NetworkX function that allows a network to be created with specific communities. The function takes in a list of $r$ values for the size of each community, as well as a symmetric $r \times r$ matrix of edge densities between the communities, as well as within each community. In our model we use two communities of size 100 and 400, respectively. The core of the network consists of 100 users and has an edge density of 0.5. The periphery consists of 400 users with an edge density of 0.01. The density of edges between the core and periphery is .025. Overall, the density of the network is .035. These densities are chosen in order to achieve a connected synthetic network approaching the density of MOOCs observed by Kellogg et al. In their study, the density of a MOOC consisting of 377 users was .01 (Kellogg et al. 2014).

In the stochastic block model, we add our UOI considering two separate scenarios:

- the UOI is connected to 10 users in the core
- the UOI is connected to 1 user in the periphery.

These scenarios look at the effect our recommender model has when a user is well established in the core, and new to the distance learning environment. This latter scenario helps to replicate and model the "cold-start" problem experienced by recommender systems.

### 3.2.2 Content Network

Our second network, the content network, is used to ensure that socially recommended chunklets are relevant to a user by determining path length to content completed by our UOI. The path length between two nodes captures the similarity between content. For strong tie recommendations, we require a short path length to UOI completed content in order to ensure applicability to a user's main interests and skills. Content recommended from weak ties on the other hand, are further from completed material in order to expose our UOI to new areas of knowledge without being so different that there is no correlation to a user's interests.

In the content network, we incorporate the current structure of the CHUNK Learning database to create a network connecting chunks and chunklets that are similar to one another. The content network graph $G_c$ contains the vertex set $V_c$ and edge set $E_c$. The vertex set $V_c$ consists of existing chunks and chunklets in the CHUNK Learning database. The edge set $E_c$ represents the relationships between any two nodes in $V_c$, namely if the respective chunks/chunklets are considered similar. We define two chunks or chunklets to be similar, and thus adjacent, if either:

- they have a parent-child relation: The parent node of a chunklet is the chunk that the particular chunklet is contained within. Thus, for chunk $u$, and chunklet $v$ contained within chunk $u$, we have edge $uv \in E_c$.
- they possess a shared identifying keyword: Currently in the CHUNK Learning database there are various types of keywords that describe characteristics of users, chunks, and chunklets. We define an "identifying keyword" as a keyword in the specific set of keyword types {"Discipline", "Topic", "Application", "Skill", "Occupational Specialty"}. If a pair of any two chunks or chunklets $u$, $v \in V_c$ shares the same identifying keyword, we then have the edge $uv \in E_c$.

### 3.2.3 User-Content Network

Our third and final network works as a conduit between the first two. Constructed as a bipartite network, this network connects users to the chunks and chunklets they have completed. Specifically, we add no connections between users, and no connections between chunks and chunklets. This network is used to determine which chunks and chunklets users of different tie strengths have completed.

The new user-content network $G_{uc}$ consists of the the vertex set $V_{uc}$, which is the union of the vertex sets from the user and content networks already introduced, i.e. $V_{uc} = V_u \bigcup V_c$. Edges exist between users and chunks/chunklets and are added as follows. Each user node ($u_i \in V_{uc}$) is randomly assigned 5 chunklets ($v_j$) that we assume $u_i$ has completed in the past, thus capturing existing knowledge that users already have obtained. The edge set $E_{uc}$ then consists of edges $\{u_iv_j, u_iv_{j+1}, u_iv_{j+2}, u_iv_{j+3}, u_iv_{j+4}\}$ for each user $u_i \in V_u$, and the five completed chunks or chunklets assigned to each user, namely $v_j, v_{j+1}, v_{j+2}, v_{j+3}, v_{j+4} \in V_c$.

## 3.3 Network Calculations

Using the social network outlined in Section 3.2.1 we compute similarity scores between users and our UOI. We use two different measures of similarity, the Jaccard Index, and the

Pearson Correlation Coefficient. The Jaccard index defined by Equation 2.1, gives insight into the amount of neighbors two nodes share, normalized by the size of their combined neighborhood. Likewise, the Pearson correlation coefficient as defined by Equation 2.2, establishes if two nodes have more or less neighbors than is expected in a strictly random network. We use the two values separately in different simulations as a way to simulate different similarity interpretations and different recommendation results.

In both cases a user who has a higher similarity score with the UOI is considered a stronger tie, while those with low similarity scores are considered weaker ties, as we identify in Section 3.4. We use this structure assuming that in a distance learning environment, users who share multiple neighbors are likely to be in the same academic field or interested in the same topics.

## 3.4 Similarity Threshold and Ordering

Once we compute the similarity score between our UOI and every user in $G_u$, we can then distinguish between users that are classified as strongly or weakly similar to the UOI. To distinguish between strong and weak ties we classify the top 20% of users in each similarity metric as a strong tie, while the bottom 80% of users are classified as weak ties. These values are consistent with the makeup of MOOCs where approximately 20% of users are in the core and 80% of users are in the periphery (Kellogg et al. 2014).

There is much discussion in academic circles on what constitutes a strong/ weak tie between two nodes in social networks. For our work, we use the definition of two users being more similar if they share more common neighbors. We expect this to be the case in social networks in that people with many common associations are more likely in the same social circle. In the context of learning environments this will suggest two individuals are in the same study group, cohort, academic department, or academic field. Under this assumption we then incorporate social capital theory to suggest that users connected by a strong tie will promote binding capital, while those connected by weak ties will promote bridging capital. Another way we refer to the benefits derived from strong ties is that they will promote depth of knowledge recommendations, while weak ties will promote breadth of knowledge recommendations.

Using the similarity scores associated with each pair of each user and our UOI from above, we classify the completed chunks and chunklets of each user as strong or weak, based on if the user is considered a strong or weak tie relative to our UOI. Chunks and chunklets can be both classified as strong or weak but will only be recommended based on our ordering restrictions explained below. Using the associated strong or weak classification of each chunk/chunklet, and the frequency of such classification, our system then recommends chunks and chunklets to our UOI according to the following the method:

Define the following:

- $G_c(V_c, E_c)$ is the content network of chunks and chunklets (for simplicity, we refer to both chunks and chunklets as chunklets below)
- $\beta_i$ is a chunklet the UOI has completed, where $i$ ranges from 1 to the total number of chunklets completed by the UOI
- $\gamma_i$ is a chunklet associated with a strong social tie user, where $i$ ranges from 1 to the total number of chunklets classified as "strong"
- $\delta_i$ is a chunklet associated with a weak social tie user, where $i$ ranges from 1 to the total number of chunklets classified as "weak."

Then:

- If $(\beta_i, \gamma_i) \in E_c$ the system recommends $\gamma_i$ as a "strong" chunklet for the UOI
- If $(\beta_i, \delta_i) \notin E_c$ the system recommends $\delta_i$ as a "weak" chunklet for the UOI.

Thus, we ensure that all strong tie recommendations are a length 1 away from content the UOI has completed in the content network $G_c$. Likewise, we require all weak tie recommendations to be a length 2 or greater away from content the UOI has completed. Since they are required to be two steps away, there will be no edge in $G_c$ between weakly recommended material and material our UOI has already completed.

Following this classification, we then separately order the results of the strong and weak recommended chunklets. For strong chunklets we order strictly based on the frequency $f$ that the chunklet is classified as strong. Thus, a chunklet completed by two strong users is ordered higher than a chunklet completed by one strong user. If two chunklets are recommended with the same frequency, the chunklet recommended from the user with the highest similarity score is recommended first. Mathematically, if chunklets $\gamma_i$ and $\gamma_j$ are recommended as strong chunklets:

$$\gamma_i \succ \gamma_j \text{ if } f(\gamma_i) > f(\gamma_j) \tag{3.1}$$

For chunklets recommended as weak, we order them first by the shortest path length ($l$) to a chunklet the UOI completed $\beta_i$ in $G_c$ and then the frequency of weak classification. Thus, recommended chunklets that are two steps away from UOI's completed material are ordered higher than chunklets that are three steps away from UOI's completed material regardless of frequency. If two chunklets are the same number of steps away from UOI's completed material, the frequency of weak classification determines their order just as in the ordering for strong chunklets. In the case where both path length and frequency are the same between two chunklets, the chunklet recommended from the user with the highest similarity score to our UOI is ordered first. Mathematically, if chunklets $\delta_i$ and $\delta_j$ are recommended as weak chunklets, then:

$$\delta_i \succ \delta_j \text{ if } \begin{cases} l(\delta_i, \beta_i) < l(\delta_j, \beta_i) \\ \text{if } l(\delta_j, \beta_i) = l(\delta_i, \beta_i) \text{ and } f(\delta_i) > f(\delta_j) \end{cases} \tag{3.2}$$

For our work, we simulate our UOI completing either 15 or 20 chunks/chunklets depending on our simulation comparison. We also assume that the UOI chooses the top recommended chunk/chunklet in all cases. We introduce a time distribution in which our UOI spends $t_S$ time on strong chunklets and $t_W$ time on weak chunklets. The time distribution is used to simulate how often our UOI will complete strong or weak recommended chunks/chunklets. We can thus tailor our simulation to account for an undecided UOI, namely a user unsure of their area of academic interest by having them focus primarily on weakly recommended material, or to account for a user who is firmly rooted in a specific subject. Following the simulation, we then determine users who have completed common chunks/chunklets with our UOI. We color these nodes and analyze the resulting recommendation sub-networks following the simulation.

THIS PAGE INTENTIONALLY LEFT BLANK

# CHAPTER 4:
# Results and Analysis

In this chapter we outline the results of our network construction and recommender model.

## 4.1 Network Construction

We begin by creating the Barabasi-Albert synthetic social network to model real social networks. Figure 4.1 shows a visualization of the BA synthetic social network. After adding the initial 500 user nodes, we add a single UOI node. We then add edges between randomly selected users in the network under two separate scenarios. In the first scenario we connect our UOI to 10 random nodes around the network, and in the second we connect the UOI to 1 random node. Doing so allows us to replicate a user that is well established in the social network, and another who is new to the network. The latter allows us to simulate the "cold-start" problem experienced by recommender models. Our methodology relies on connection to other users, and thus a disconnected UOI results in no recommendations. We note that the modularity value of our BA graph is 0.19 with a total of 8 communities according to the Louvain algorithm.

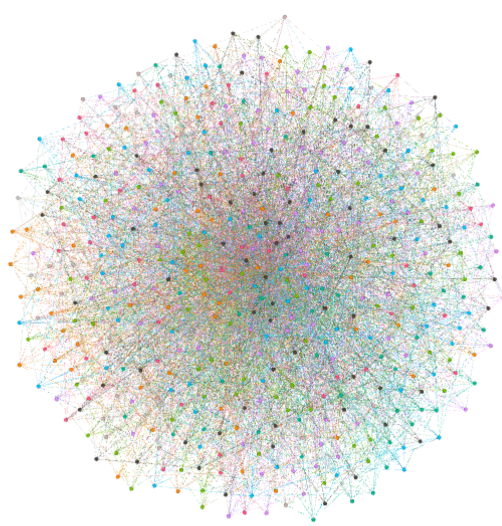

Figure 4.1. Barabasi-Albert synthetic social network consisting of 500 user nodes and one UOI node. Node colors are determined by modularity class.

Similarly, our stochastic block model synthetic social network, presented in Figure 4.2, simulates the structure of a social network modelling a core-periphery structure. For our two scenarios in the SBM, we add our UOI node after creating the SBM of 500 users. Then by looking at the degree distribution of the 500 nodes, we found that nodes in the core all had degree greater than 43, while nodes in the periphery had much smaller degree. Separating the nodes based on degree, we then were able to randomly assign edges between our UOI to 10 users in the core, and 1 user in the periphery. As mentioned previously, these scenarios replicate a user who is well established in the core of our network, and new to the learning environment. We note that the modularity value of our SBM is 0.189 with 6 total communities.

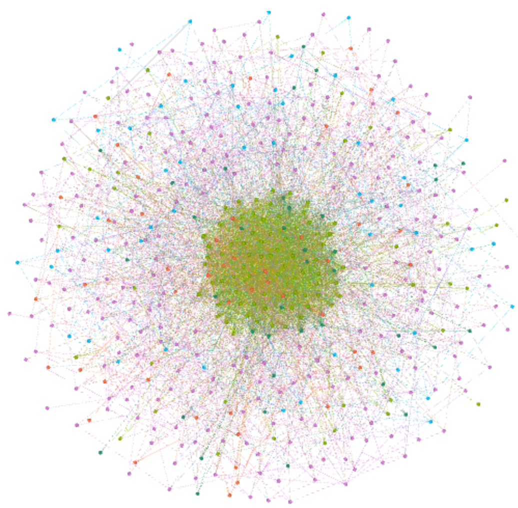

Figure 4.2. Stochastic block model social network consisting of 500 user nodes and one UOI node. The core contains 100 tightly connected users while the periphery consists of 400 sparsely connected users. Node colors are determined by modularity class.

The next network we use in our recommender model is the content network of material in the CHUNK Learning database. Each node represents a chunk or chunklet with edges between those that share common keywords or possess a parent-child relationship. Figure 4.3 displays the current CHUNK network as of January 2021. We color each node based on its modularity class.

In Figure 4.3, we visually observe about 6 strongly outlined modularity classes. These different modularity classes generally separate topics taught in CHUNK Learning. For the

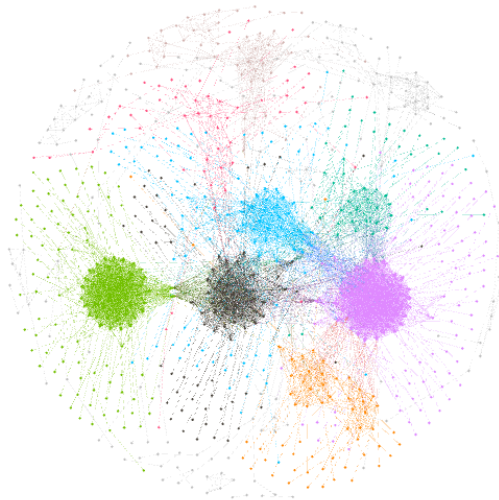

Figure 4.3. Content network of the CHUNK Learning database colored by modularity class

purposes of our simulations we make special note of the community of black nodes depicted in the middle of Figure 4.3. This community of 89 nodes, represents the "Fundamentals of Physics" topic taught in CHUNK Learning. We will use a physics-oriented student to compare our recommendation results to that of previous models.

The content network in Figure 4.3 has a modularity value of .717 suggesting that there are strong communities in the network. We can observe three very strong communities in the middle of the network, with surrounding branching communities. In all however, we calculate there to be 29 separate modularity classes in the entire network.

We note that several topics currently taught in CHUNK Learning do not appear as large obvious communities. One example is a course titled "Structure and Analysis of Complex Networks". Chunks and chunklets involved with this topic are located in the periphery of the network. We attribute this to how material is tagged within CHUNK. Content in our three tightly clustered communities representing the "Nuclear Deterrence", "Fundamentals of Mathematics", and "Fundamentals of Physics" topics more consistently possess unique identifying keywords. We note that chunks and chunklets within "Structure and Analysis of Complex Networks" are tagged both less frequently and with less identifying keywords than the three strong communities we observe in Figure 4.3. These weakly outlined communities are thus less integrated in the network or connected to a wider range of other chunks/chunklets from different topics. We make special note that our methodology relies on effective tagging practices for material in order to take advantage of the community structure we observe from “Nuclear Deterrence”, “Fundamentals of Mathematics”, and “Fundamentals of Physics”.

## 4.2 Recommender Model Results on Random UOI

We conduct the first test of our recommender model with a UOI who has completed five randomly assigned chunks/chunklets. We then simulate the completion of 20 additional chunks/chunklets over the course of a semester. We test our recommender system when it recommends material based solely on strong social ties and separately on solely weak social ties. Figure 4.4 and Figure 4.5 display the results of our exclusively strong tie simulations.

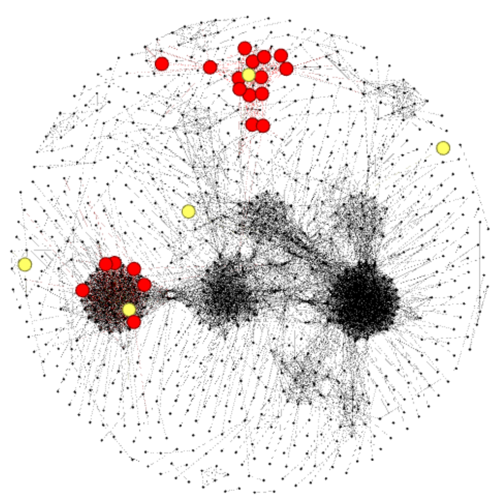

Figure 4.4. UOI with 5 randomly assigned chunks/chunklets, connected to 10 random users in a BA synthetic network. Content recommendations made based on strong social ties as determined by the pairwise Jaccard index. Nodes in yellow are the 5 initial chunklets completed by the UOI and red nodes are recommended content.

For results in Figure 4.4 we use the BA synthetic network model where our UOI is connected to 10 users, and similarity scores are calculated using the Jaccard index. For the results displayed in Figure 4.5 we use the BA synthetic social network model where our UOI is connected to 10 users, and similarity scores are calculated using the Pearson correlation coefficient. In both cases, we observe tight clustering around a previously completed

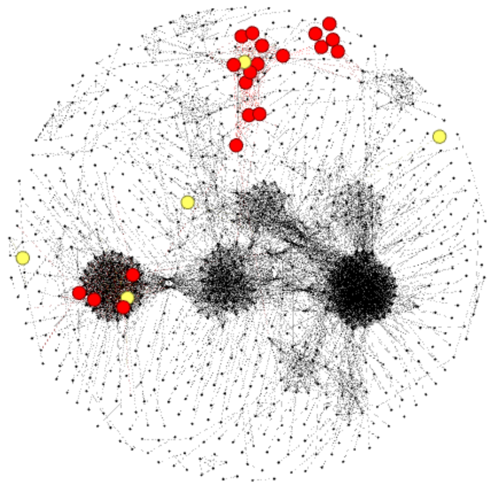

Figure 4.5. UOI with 5 randomly assigned chunks/chunklets, connected to 10 random users in a BA synthetic network. Content recommendations made based on strong social ties as determined by the pairwise Pearson correlation coefficient. Nodes in yellow are the 5 initial chunklets completed by the UOI and red nodes are recommended content.

chunk/chunklet. Our simulation began with our UOI attached to the the five yellow nodes. We finish the simulation with five communities centered on initially completed content. This demonstrates our condition that the most highly recommended material from strong user ties are directly related to a user's field of interest. We note that in both instances the recommended chunks/chunklets are different due to the different similarity metrics used to determine strong vs. weak social ties.

Similarly, we simulated our user completing 20 chunks/chunklets starting with the same five randomly assigned chunks/chunklets but only recommending material based on weak social ties. The results of these two simulations are outlined in Figure 4.6 and Figure 4.7.

In Figure 4.6, we use the SBM synthetic social network with our UOI connected to 10 users in the core, and similarity scores calculated using the Jaccard index. In Figure 4.7, we use the SBM synthetic social network with our UOI connected to 1 user in the periphery, and similarity scores calculated using the Jaccard index. Unlike in Figure 4.4 and Figure 4.5, we now see a network of recommended chunks/chunklets that are disconnected from one another. There are no direct edges between any of the chunks/chunklets the UOI completed. We observe no community structure in this network and achieve recommending our UOI

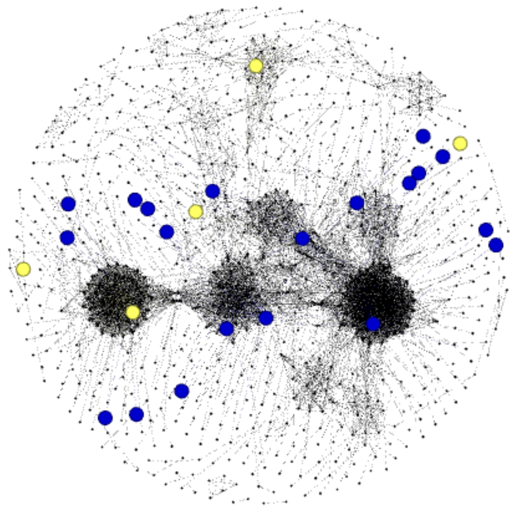

Figure 4.6. UOI with randomly assigned content completion connected to 10 random users in the core of our SBM. Content recommendations based solely on weak social ties as determined by the pairwise Jaccard index. Nodes in yellow are the 5 initial chunklets completed by the UOI.

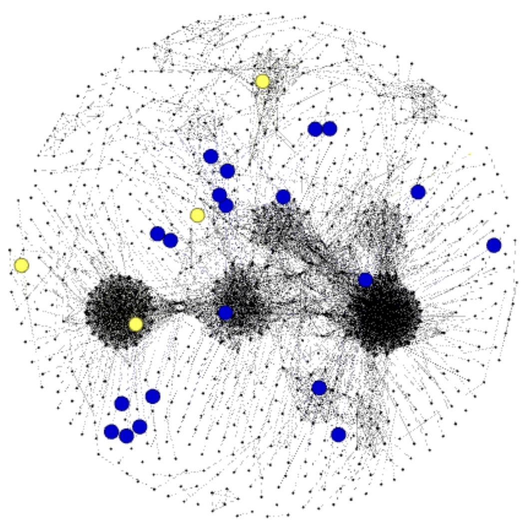

Figure 4.7. UOI with randomly assigned content completion connected to 1 random users in the periphery of our SBM. Content recommendations based solely on weak social ties as determined by the pairwise Jaccard index. Nodes in yellow are the 5 initial chunklets completed by the UOI.

a wide breadth of applicable material tangentially related to their previously completed material.

Below in Table 4.1 we outline the network statistics for the four previous strong or weak simulations. We see that in the cases of strong social tie recommendations we have greater density and average clustering coefficient than the entire content network. For our strong tie recommendations, we see a completed chunk/chunklet sub-network density of 0.473 and 0.393 while the full content network has a density of 0.024. Likewise, our average clustering coefficient for the completed chunk/chunklet sub-network is 0.929 and 0.903, whereas for the entire content network the value is 0.666. In the case of the weak chunk/chunklet sub-networks we are left with a disconnected network with zero edges between any two nodes and 25 separate communities. This is exactly what we desire from our weak social tie recommendations, by allowing our UOI to access a wider breadth of knowledge that is tangentially related to the chunks/chunklets they have completed previously. The two cases of only strong and only weak ties are the two extrema of our recommendation system and we demonstrate in the next section a more realistic distribution.

Table 4.1. Recommendation Sub-network Statistics of Strong and Weak Simulations

| Metric | Figure 4.4 | Figure 4.5 | Figure 4.6 | Figure 4.7 |
|---|---|---|---|---|
| Graph Density | 0.473 | 0.393 | 0 | 0 |
| Avg. Node Degree | 11.36 | 9.44 | 0 | 0 |
| Avg. Clustering Coef. | 0.929 | 0.903 | N/A | N/A |
| Modularity | 0.274 | 0.462 | N/A | N/A |
| # of Communities | 5 | 5 | 25 | 25 |
| # of Edges | 142 | 118 | 0 | 0 |
| % of Edges | 1.72% | 1.43% | 0% | 0% |

## 4.3 Recommender Model Results vs. Existing Work

In the next test of our recommender model, we compare the recommendation results of our methodology to that of Diaz et al. (2019). In their work, Diaz et al. developed a recommender model for CHUNK Learning based on dynamic keyword attachment. One of the simulations of their work focused on a user who was interested in physics. This user's profile was initially given keywords "rocket," "physics," "newton," and "motion".

To compare our results, we simulate our UOI beginning with similar initial conditions. Te replicate the initially assigned keywords, our simulation requires a chunk/chunklet to possess all the same keywords. Since there were no combination of chunks/chunklets that possessed the same four keywords, we began with another chunklet that possessed four basic physics keywords. We started our UOI on a chunklet titled "Simple Harmonic Motion - Graphs of Position, Velocity, and Acceleration." The attached keywords were "physics", "position", "velocity", and "acceleration". For comparison, we plot the results of Diaz et al. in the content network in Figure 4.8.

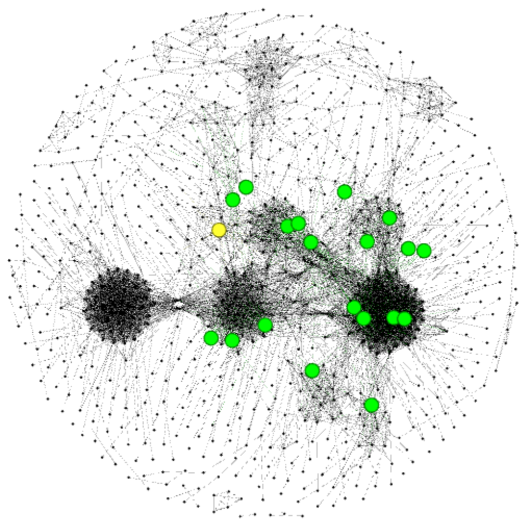

Figure 4.8. A visualization of the content network and recommendations from (Diaz et al. 2019). Their recommender model focuses on dynamic user profile keyword attachment for a student interested in Physics. Node in yellow is the starting chunklet "Vectors," and green nodes are highest recommended nodes after each previous chunk/chunklet completion.

In Figures 4.9-4.12 we display the results of our simulations. Figure 4.9 depicts the results for a UOI in the BA synthetic social network connected to ten users. Figure 4.10 is the simulation result of a UOI in the BA synthetic social network connected to one user. Figure 4.11 shows the results of a UOI in the SBM synthetic social network connected to one user in the periphery. Figure 4.12 shows the results of a UOI in the SBM synthetic social network connected to ten users in the core.

In our simulations we see a generally tight grouping of completed red chunks/chunklets. These recommended chunks/chunklets arise from strong social ties and provide the UOI

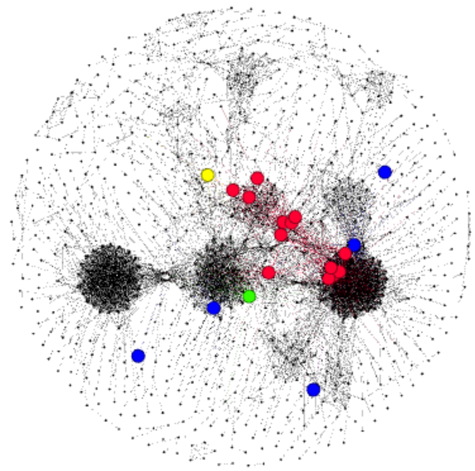

Figure 4.9. Post simulation for a user focused on Physics using the BA social network, where our UOI is connected to 10 other users in a social network. The Jaccard index is used to determine social tie strength. The yellow node corresponds to the "Vectors" chunklet and the green node represents the initial chunklet our UOI completed to simulate the attachment of keywords as done by Diaz et al. Red nodes are chunks/ chunklets recommended as "strong" and blue nodes are recommended as "weak."

with deeper instruction on their topic of interest. We see in Figure 4.11, and Figure 4.12 that the red nodes are all located within and around the "Fundamentals of Physics" community we highlighted earlier. In Figure 4.9 and Figure 4.10 we see a looser grouping of red nodes that are contained within the "Fundamentals of Mathematics" community.

Although related to physics, we attribute the shift away from the physics community to our simulation parameters. Our simulation is run such that the top recommended chunk/chunklet is considered completed after the simulation step. At the first iteration of recommendation and simulation, there is a 25% chance that the recommendation is from a weak tie and separate from our initial chunks/chunklets, or a 75% chance that the recommendation is taken from a neighbor of either the "Simple Harmonic Motion" or "Vectors" chunklets. The neighbor is equally likely to come from "Simple Harmonic Motion" or "Vectors." At the next iteration there was then an equal chance the next recommended chunklet came from the "Simple Harmonic Motion," "Vectors" or the first completed chunk/chunklet and again a 25% chance the recommendation comes from a weak tie and is separate.

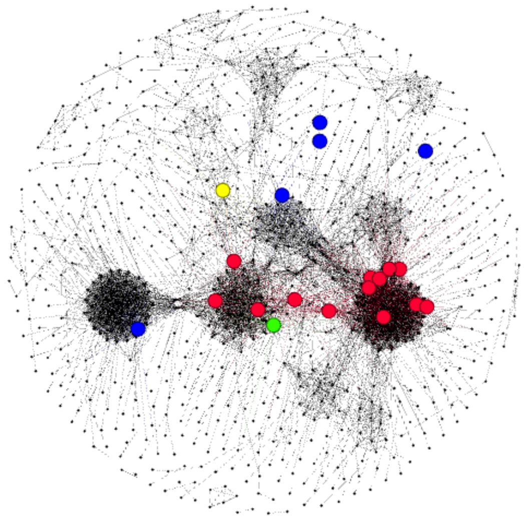

Figure 4.10. Post simulation for a user focused on Physics using the BA social network, where our UOI is connected to 1 other user in the social network. The Jaccard index is used to determine social tie strength. The yellow node corresponds to the "Vectors" chunklet and the green node represents the initial chunklet our UOI completed to simulate the attachment of keywords as done by Diaz et al. Red nodes are chunks/ chunklets recommended as "strong" and blue nodes are recommended as "weak."

With these processes repeating over and over each time, our simulation can have a shift in the community recommendations are centered around. In Figure 4.9 and Figure 4.10 the recommended content more often came from material in the mathematics community. This can be viewed as an advantage of our recommender system by recognizing and adapting to our UOI's preferences. In this case the UOI would have shifted from physics to mathematics, and our recommender model aided them in that transfer.

Comparing our results to Diaz et al. we see that there is improvement in the density of recommended material. Looking at Table 4.2 we see that the average density of the four simulations above is 0.5 while the density of recommendations from Diaz et al. is 0.058. Our densities for the simulations above range from $0.358 to 0.616$, thus for each of our simulations, our recommender system demonstrates a significant improvement. Under the construction of our content network the density of the sub-network is a measure of how associated the recommended chunks/chunklets are. We thus are able to provide improved accuracy in the relevancy of our recommendations.

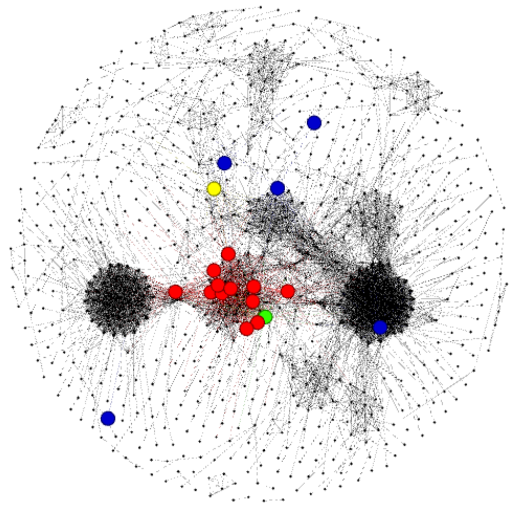

Figure 4.11. Post simulation for a user focused on Physics using the SBM social network, where our UOI is connected to 1 other user located in the periphery of the network. Pearson correlation coefficient is used to determine social tie strength. The yellow node corresponds to the “Vectors” chunklet and the green node represents the initial chunklet our UOI completed to simulate the attachment of keywords as done by Diaz et al. Red nodes are chunks/ chunklets recommended as “strong” and blue nodes are recommended as “weak.”

Furthermore, our recommendation sub-network has greater average node degree, greater number of total edges, and a larger clustering coefficient than the sub-network recommended by Diaz et al. The average degree of our sub-networks ranged from 6.8 – 11.7, the total number of edges in the sub-network ranged from 68 – 117, and the average clustering coefficient ranged from 0.836 – 0.99. These values, like density above, correspond to a deeper connection between recommended material.

We also note that our average modularity value is lower than the results of Diaz et al., suggesting that there are less sub-communities within our recommendations. This again supports our conclusion that our recommendations are more uniform than recommendations from dynamic keyword attachment. We expect the modularity value would increase with an increase in the number of weak tie chunk/chunklet completed. With each new weak tie chunk/chunklet we introduce a new topic, and therefore new community, to our recommended sub-network. We also note that we have an average of 7 communities in our

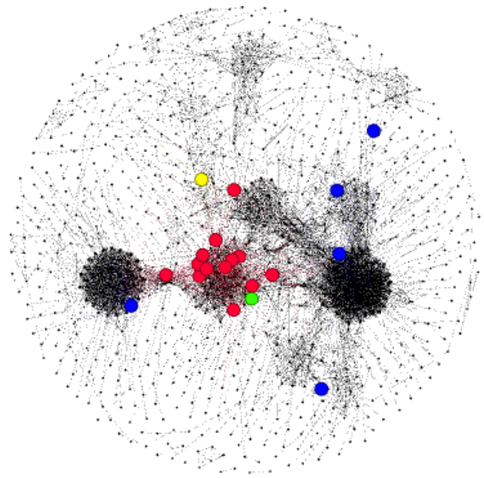

Figure 4.12. Post simulation for a user focused on Physics using the SBM social network, where our UOI is connected to 10 users located in the core of the network. Pearson correlation coefficient is used to determine social tie strength. The yellow node corresponds to the "Vectors" chunklet and the green node represents the initial chunklet our UOI completed to simulate the attachment of keywords as done by Diaz et al. Red nodes are chunks/chunklets recommended as "strong" and blue nodes are recommended as "weak."

simulations which corresponds to an individual community for the five weak tie recommendations and 2 communities on average for our strong tie recommendations.

Table 4.2. Recommendation Sub-Network Statistics

| Metric | (Diaz et al. 2019) | Avg. of our Simulations |
|---|---|---|
| Graph Density | 0.058 | 0.5 |
| Avg. Node Degree | 1.1 | 9.5 |
| Avg. Clustering Coef. | 0.833 | 0.9175 |
| Modularity | 0.45 | 0.166 |
| # of Communities | 14 | 7 |
| # of Edges | 11 | 91.75 |
| % of Edges | 0.13% | 1.11% |

We highlight the existence of blue nodes spread widely throughout the network as an important aspect of our recommender model. These chunks/chunklets, recommended from weak social ties, provide our UOI with an educational asset by exposing them to material they otherwise would have less access too. Relating back to the theory of connectivism, our recommender system will encourage the UOI to learn new ways of thinking and apply it to their curriculum of interest or share their skills in a new field. In either case the additional asset provided by our recommender system aids students in adapting and succeeding in the 21st century world.

## 4.4 Social Network Development

Under our assumptions, we associate content and users randomly throughout the network. Thus, of the five chunks/chunklets a user completes, there is high likelihood that they are disconnected within the content network. This is not very realistic, and we discuss it further in future work. However, through our completely synthetic social networks and randomized content association we can still begin to see the social benefits of leveraging social ties in the recommender model. Figure 4.13 shows the UOI's social network before a simulation.

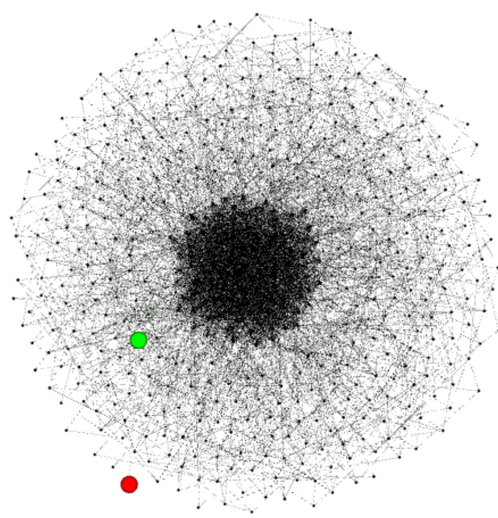

Figure 4.13. Social Network of UOI before simulation. Social network is the SBM where our UOI is connected to 1 random user in the periphery of the network. UOI is red node.

The network in Figure 4.13 is the simulation of our UOI who is interested in physics, and is connected to 1 other user in the periphery of a SBM. In Figure 4.14 we see the users who have completed at least one common chunk/chunklet as our UOI.

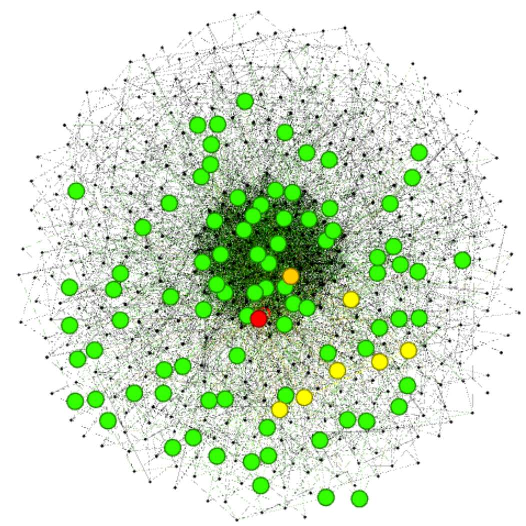

Figure 4.14. Social network of UOI after simulation. UOI in red. Nodes in green represent users who have completed 1 common chunks/chunklet with the UOI. Nodes in yellow represent users who have completed 2 common chunks/chunklets with the UOI.

Immediately we see the much larger integration of our UOI in the network. The degree of our UOI node has increased from 1 to 95 following the simulation of 20 chunks/chunklets outlined in Figure 4.11. More importantly however, we observe the formation of a clique between our nodes in yellow.

The nodes in yellow are nodes who have completed two common chunks/chunklets with the UOI. Given that our UOI is interested in physics and recommended primarily content within the physics topic we see that the developing clique can be considered a physics cohort of the larger social network. We make note of the relative positioning of users in this clique by their close proximity. Our network is displayed using the Fruchterman-Reingold layout in Gephi, which simulates edges between nodes as springs and seeks to minimize the “energy” of the system. This means that the close proximity between each yellow and red node suggests a strong relation among the group. We expect that for simulations on multiple users and analysis on real data, this proximity would increase, and sub-network density of the clique would suggest a strongly connected community. It is through this algorithmic formation of communities that we believe our recommendation system will connect students who would benefit from an academic relationship. Providing users this information can encourage them to discuss ideas with peers they are sure are interested in similar topics. From this general foundation, we believe that overtime, a community of learners will develop, ideally culminating in the formation of a community of practice as outlined in Section 2.2.2.

THIS PAGE INTENTIONALLY LEFT BLANK

# CHAPTER 5: Conclusions and Future Direction

In this chapter we summarize what we have learned in this research, and propose further extensions of our work.

## 5.1 Conclusions

Our initial goal was to address and improve upon the lack of socialization in distance learning (DL) environments. As constructed, most DL environments are solitary in nature. In this work we developed a recommender system that would not only improve recommendations for students in CHUNK Learning, but that would also promote the formation of communities of learners with similar interests. These communities will provide academic support for users and will allow individuals to expand their network in academia. Overtime, these communities may develop high levels of cohesion and develop into communities of practice within a DL environment.

To improve content recommendations to users in a DL environment, we utilized social capital theory to leverage the types of ties individuals have in their social networks to guide their education. Strong ties correspond to close supportive relationships between people. In a learning environment we take these ties to be among peers who share the same academic interest or are in the same discipline, department, class, or cohort. This will provide a better opportunity to identify individuals with similar knowledge bases, and thus will help to cement knowledge in a common academic field. Social capital theory refers to the benefit derived from strong ties as bonding capital, and in learning environments we refer to the academic benefit as improved depth of knowledge. Weak ties on the other hand correspond to more distant relationships between people. These are relationships between individuals who interact less often, or not at all. In a DL environment, we can consider these ties between peers of differing academic departments. Social capital theory refers to the benefit derived from weak ties as bridging capital. In the learning environment we refer to the academic benefit as an improved breadth of knowledge, since more distant individuals likely have a different area of expertise and can expose individuals to new material and ways of thinking.

Using the benefits derived from strong and weak ties, our system recommends content that is more precise, accurate, and better tailored to a student's preferences than other recommender models. Our results suggest a higher density, average degree, and clustering coefficient in the recommended content sub-network. The combination of improved depth and breadth of knowledge helps to encourage learning by connecting new knowledge to an individual's wider knowledge base and strengthening the connection between what an individual already knows. Using the learning theory of connectivism, we believe such a recommendation system can improve learning outcomes for students.

Furthermore, by using an individual's network to drive their learning, we facilitate the formation of communities among users studying similar material. Although there is no guarantee that connecting users through content will ensure heightened user engagement, we believe that by connecting users completing similar material online we are able to give users the information to improve socialization even when studying in disparate locations. The algorithmic formation of such communities within a DL environment empowers students to take control of their academic network and learn from peers as they deem fit.

## 5.2 Further Direction

To improve upon our work, we suggest the following areas to focus on: community formation, recommendations, and computational efficiency. We recognize this list is not exhaustive.

To improve community formation, we note the need for further research and investigation into what a social network looks like in a learning environment. This would require real user data at a minimum. It could also be expanded to incorporate user input for connections, and could include algorithms for computing similarity among individuals based on keyword attachment.

For recommendations, a direct improvement can occur in how we order content. In this paper we used similarity metrics, distance between content in the network and frequency of classification. Our recommendations may be improved by incorporating similarity based metrics between content such as variants of cosine similarity. Additionally, we can incorporate weighted edges in the content network to connect content more efficiently. This would allow the system to better tailor recommendations strongly associated with a specific content a user enjoys and minimize or maximize the amount of association for weak recommendations.

Finally, we note that the computational complexity of the algorithm will grow exponentially for a learning environment where recommendations are made for large numbers of users. In addition to recommendations being accurate, precise, and tailored to user preferences, they also must be timely and efficiently calculated. We thus note that a future step in implementing our system is to apply the recommendations at scale and investigate ways that improvements in computational efficiency can be achieved.

THIS PAGE INTENTIONALLY LEFT BLANK

# List of References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74(1):47.

Aleta A, Moreno Y (2019) Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics* 10:45–62.

Barabasi AL (2016) *Network Science* (Cambridge University Press, Cambridge, UK).

Bianconi G (2018) *Multilayer Networks: Structure and Function* (Oxford University Press, Oxford, UK).

Bollobás B (1998) *Modern Graph Theory* (Springer, New York), doi:10.1007/978-1-4612-0619-4.

Causey J, Harnack-Eber A, Ryu A, Shapiro D (2021) A covid-19 special analysis update for high school benchmarks. Technical report, Herndon, VA: National Student Clearinghouse Research Center.

Cohen WM, Levinthal DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35(1):128–152.

Diaz DO, Gera R, Keeley PC, Miller MT, Leondaridis-Mena N (2019) A recommender model for the personalized adaptive chunk learning system. Monterey, California. Naval Postgraduate School.

Gera R, Bartolf D, Isenhour M (2020) Chunk learning. URL http://chunklearning.net/.

Grunspan DZ, Wiggins BL, Goodreau SM (2014) Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE—Life Sciences Education* 13(2):167–178.

Haythornthwaite C (2019) Learning, connectivity and networks. *Information and Learning Sciences* 120(1/2):19–38.

Haythornthwaite C, De Laat M (2010) Social networks and learning networks: Using social network perspectives to understand social learning. *Proceedings of the 7th international conference on networked learning*, 183–190 (Aalborg University, Aalborg, DK).

Kellogg S, Booth S, Oliver K (2014) A social network perspective on peer supported learning in moocs for educators. *International Review of Research in Open and Distributed Learning* 15(5):263–289.

Lai CH, Lin HW, Lin RM, Tho PD (2019) Effect of peer interaction among online learning community on learning engagement and achievement. *International Journal of Distance Education Technologies (IJDET)* 17(1):66–77.

Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. *Decision Support Systems* 74:12–32.

Newman ME (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23):8577–8582.

Newman MEJ (2018) *Networks: An Introduction,* 2nd ed. (Oxford University Press, Oxford, UK).

Orman GK, Labatut V, Cherifi H (2013) Towards realistic artificial benchmark for community detection algorithms evaluation. *International Journal of Web Based Communities* 9(3):349–370.

Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D (2004) Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences* 101(9):2658–2663.

Ricci F, Rokach L, Shapira B (2015) Introduction to recommender systems handbook. *Recommender Systems Handbook*, 1–34 (Springer, New York).

Rombach P, Porter MA, Fowler JH, Mucha PJ (2017) Core-periphery structure in networks (revisited). *SIAM Review* 59(3):619–646.

Russo TC, Koesten J (2005) Prestige, centrality, and learning: A social network analysis of an online class. *Communication Education* 54(3):254–261.

Siemens G (2005) Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning* 2(1):3–10.

Smith SU, Hayes S, Shea P (2017) A critical review of the use of wenger's community of practice (cop) theoretical framework in online and blended learning research, 2000-2014. *Online Learning* 21(1):209–237.

UNICEF (2020) Education and covid-19. UNICEF Data, New York, accessed: 2021-06-04.

Venter A (2019) Social media and social capital in online learning. *South African Journal of Higher Education* 33(3):241–257.

Wang X, Lu W, Ester M, Wang C, Chen C (2016) Social recommendation with strong and weak ties. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 5–14.

West DB, et al. (2001) *Introduction to graph theory*, volume 2 (Prentice Hall, Upper Saddle River, NJ).

THIS PAGE INTENTIONALLY LEFT BLANK

# Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California