Spring 2018

Application of Evolutionary Network Concept in Structuring Mathematics Curriculum

Aditi Mitra

APPLICATION OF EVOLUTIONARY NETWORK CONCEPT IN STRUCTURING MATHEMATICS CURRICULUM

by

ADITI MITRA

(Under the Direction of Hua Wang)

ABSTRACT

Phylogenetic tree, and in general, evolutionary network, has found its application well beyond the biological fields and has even percolated into recent high demanding areas, such as data mining, and social media chain reactions. An extensive survey of its current applications are presented here. An attempt has been made to apply the very concept in the mathematics course curriculum within a degree program. Main features of the tree structure are identified within the curriculum network. To highlight various key components and to enhance the visual effect, several diagrams are presented. The combined effect of these diagrams highlights a curriculum tree structure. The current study can be used as a potential tool for effective student advisement, student placement within the curriculum, efficient resource allocation, etc. Future work may encompass detailing and implementing these applications.

INDEX WORDS: Evolutionary network, Mathematics curriculum, Georgia Southern University

2009 Mathematics Subject Classification: 15A15, 41A10

APPLICATION OF EVOLUTIONARY NETWORK CONCEPT IN STRUCTURING

MATHEMATICS CURRICULUM

by

ADITI MITRA

B.Sc., Calcutta University, India, 1988

B.S., University of Nevada, Reno, 1995

C.S., University of Nevada, Reno, 1999

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial

Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

APPLICATION OF EVOLUTIONARY NETWORK CONCEPT IN STRUCTURING MATHEMATICS CURRICULUM

by

ADITI MITRA

Major Professor: Hua Wang
Committee: Goran Lesaja
Ha Nguyen
Hua Wang

Electronic Version Approved:
May 2018

DEDICATION

This thesis is dedicated to my late father Mr. Asoke Kumar Sen.

ACKNOWLEDGMENTS

I wish to acknowledge Dr. Hua Wang for his valuable advice and guidance, as my primary thesis advisor. I would also like to acknowledge Dr. Goran Lesaga and Dr. Ha Nguyen, other thesis committee members, for their willingness to serve in my thesis committee and reviewing my work. I would also like to thank my husband, Dr. Aniruddha Mitra for his encouragement and help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF SYMBOLS

A symbol table can be created in various ways. Here are a few:

Tabular environment:

$\mathbb{R}$	Real Numbers
$\mathbb{C}$	Real Numbers
$\mathbb{Z}$	Integers
$\mathbb{N}$	Natural Numbers
$\mathbb{N}_0$	Natural Numbers including 0
$L_p(\mathbb{R})$	p-integrable functions over $\mathbb{R}$
$L(X,Y)$	Linear maps from X to Y
$\mathrm{rank}(T)$	Rank of a linear map

Multicols environment:

$\mathbb{R}$ Real Numbers

$\mathbb{C}$ Real Numbers

$\mathbb{Z}$ Integers

$\mathbb{N}$ Natural Numbers

$\mathbb{N}_0$ Natural Numbers including 0

$L_p(\mathbb{R})$ p-integrable functions over $\mathbb{R}$

$L(X,Y)$ Linear maps from X to Y

$\mathrm{rank}(T)$ Rank of a linear map

Itemize environment:

- $\mathbb{R}$ Real Numbers
- $\mathbb{C}$ Real Numbers

- $\mathbb{Z}$ Integers
- $\mathbb{N}$ Natural Numbers
- $\mathbb{N}_0$ Natural Numbers including 0
- $L_p(\mathbb{R})$ p-integrable functions over $\mathbb{R}$
- $L(X, Y)$ Linear maps from X to Y
- $\mathrm{rank}(T)$ Rank of a linear map

CHAPTER 1

INTRODUCTION

1.1 Background information

Simply put, the term Phylogeny refers to the study of evolution. Given a set of points, most of the time current living species, a phylogenetic tree refers to the “tree of life” that describes the evolutionary history of all life on Earth. The ultimate goal of any phylogenetic analysis is to reconstruct some part of the tree of life, based on the information from current species. Very often these points corresponding to the current species are called taxa corresponding to the leaves of the phylogenetic tree.

Two species are related, if they share a recent common ancestor. The species A is more closely related to B than to C, if the last common ancestor of A and B is more recent than the last common ancestor of A and C. As an example, humans are more closely related to mice (last common ancestor about 100 million years ago), than they are to fruit-flies (last common ancestor 600 million years ago).

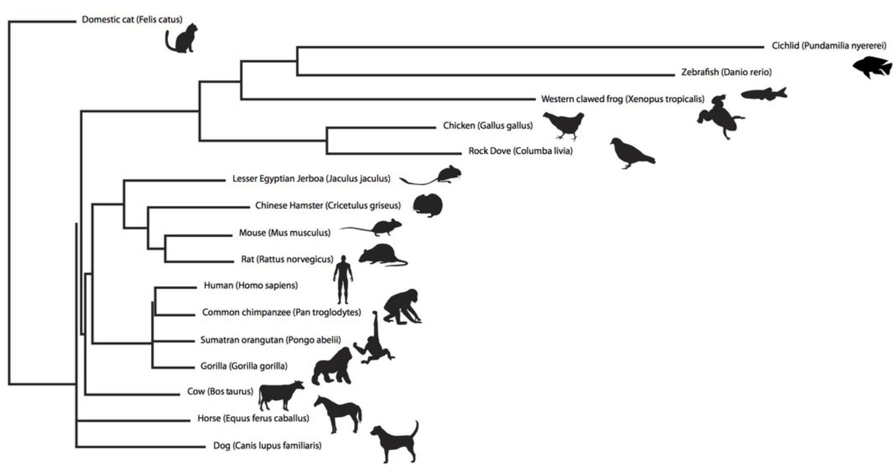

Figure 1.1: A phylogenetic tree [60].

Following the motivation of phylogenetic analysis the obvious application of this concept is to, as part of the effort to find out the evolutionary history of your taxa, examine how different taxa are related to each other.

In addition to the above mentioned basic information, there are many other ways, both practical and pure, to use the concept of phylogeny structures.

- Phylogenetic diversity can be measured using phylogenetic trees in conservation biology (and ecology in general). For this purpose the branches on the evolutionary tree are drawn to a specific length that was calculated. In general long branches usually mean that you have basal taxa that are originated earlier, and hence have more information on the evolution, and consequently more important to conserve. On the other hand, very short branches generally implies very new species, possibly the result of an adaptive radiation.

- Phylogenetic trees are also used in the search for natural products. For example, a sponge species that produces a useful chemical but also has a serious side effect in any drug produced from it. It is natural, then, to survey the sponge genus and find the species most related to it. One can then try out those species to produce similar drug in order to avoid the side effect.

- In medicine, phylogenetic trees are used to study infectious bacteria and viruses to trace their evolutionary histories. By examining what trends they have undergone in their history, one may gain valuable insights into how they have gotten canalised (i.e. restricted in their biology). If the cell wall contains a very characteristic and unchanging motif, then that is a potential good target, and you can only see such conserved motifs using a phylogenetic tree.

- Instead of tracing the evolutionary history of specific species, one may also use phylogenetic trees to guide our search for new species. This is usually done by plotting the phylogenetic tree on a physical map. Similarly for fossil searches, phylogenetic trees can be used to record when taxa originated and where.

1.2 Some formal definitions and tools

In this section we provide some formal definitions and concepts that are useful in our work later.

A set X is a set of taxa in which each taxon x_i represents some species, group, or individual organism. For example, $X = \{x_1, x_2, \ldots\}$ might denote a set of mammals, where x_1 represents gorillas, x_2 represents seals etc.

A cluster is any subset of X, excluding empty set 0 and the full set X. The topic of clustering concerns putting similar data points (or taxa, in this case) in the same subgroups, which is exactly what a phylogenetic tree shows us. Naturally, similar species are closer to each other in the evolution process, and are grouped together like the above mentioned mammals.

Sequence alignments

One of the more specific tools/concepts used in comparing species is known as "sequence alignments". This is to compare biomolecular sequences, as one of the most fundamental operations in computational biology (DNA etc). In such studies two DNA or protein sequences that have a high similarity are called Homologous (i.e., two have evolved from a common ancestral sequence). To illustrate such an idea one usually write one sequence above the other and try to maximize the number of similar or identical bases or amino acids (build proteins).

Spanning tree

Another interesting concept is the spanning trees, formally defined as an acyclic subgraph that contains all nodes. Intuitively speaking, a spanning tree of a graph or network is the most efficient way to connect all nodes using existing links from the graph/network.

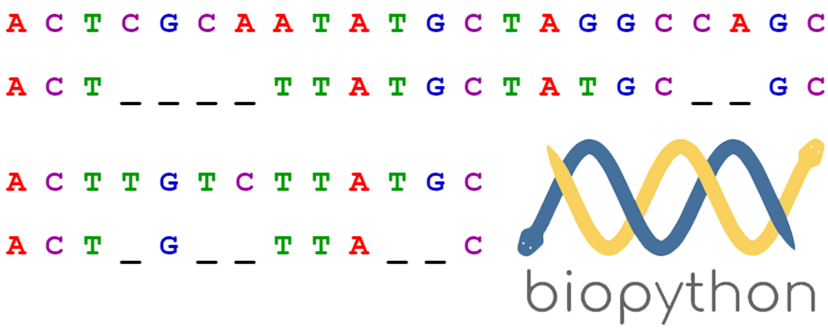

Figure 1.2: sequence alignment [62].

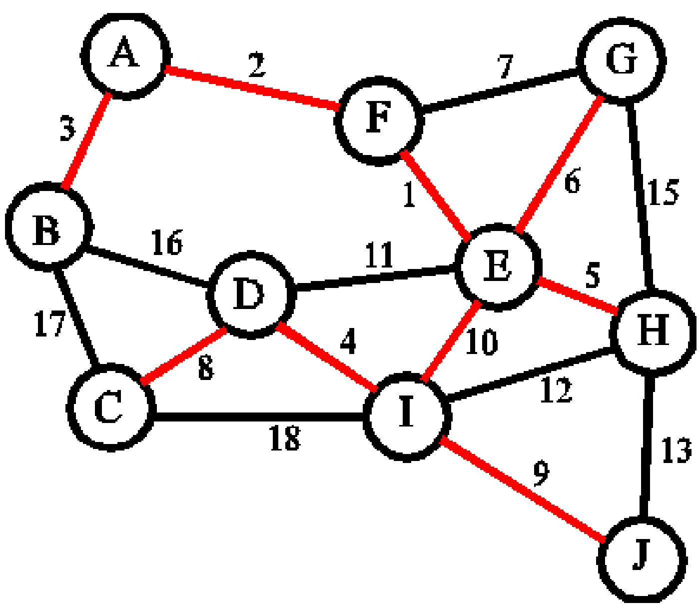

Figure 1.3: A spanning tree [61].

A single graph can have many different spanning trees. When the edges or links are weighted according to specific applications, it has been of interest to find the spanning tree with minimum total weight. This is, intuitively, still consistent with the idea of finding most efficient way of connecting a given graph/network.

A Minimum Spanning tree (MST) of an edge-weighted graph is a spanning tree whose weight (the sum of the weights of its edges) is less than or equal to the weight of every other spanning tree.

1.3 Evolutionary networks

The phylogenetic tree is a form of evolutionary trees that is "acyclic", i.e. there is no cycle. This is under the assumption that every species can only split into two or more species, but no two or more species can ever "merge" into one. In many practical applications, however, that is not the case, which leads to so called evolutionary networks, that are essentially directed graphs in which each node represent some data point and the directed edges between nodes represent the direction of evolution and the relations between nodes.

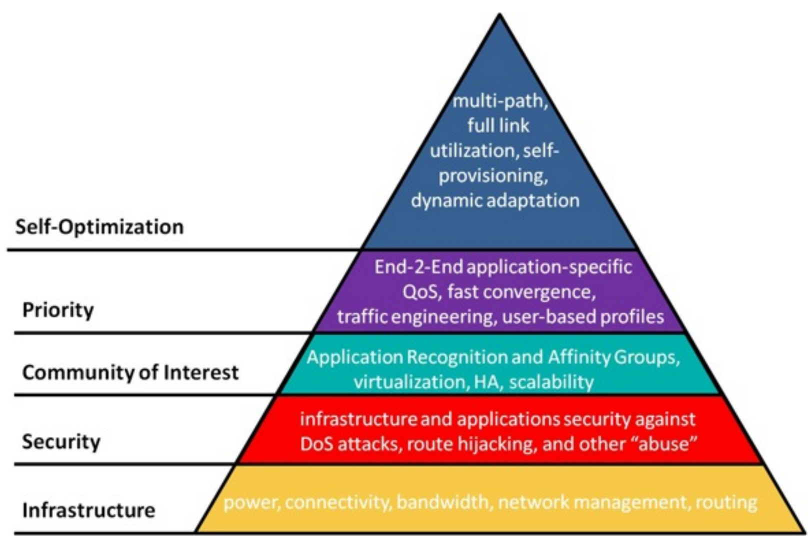

Figure 1.4: Evolutionary Network [63].

Evolutionary network analysis has become an important field of study due to the importance of different kinds of social networks, email networks, biological networks, etc. It is important to keep up with the changes of any particular network, as all topological properties and other characteristics develop according to these changes.

Such analysis of networks in general is challenging, and it requires much knowledge and manipulation of the aforementioned concepts and tools.

1.4 Application to mathematics curriculum

In this thesis we will consider different college mathematics classes as nodes and examine their relations in terms of prerequisites. For the purpose of the current study, we ignore the Developmental Mathematics classes that are offered at Georgia Southern University. A directed edge is established if the starting node is a direct prerequisite of the ending node. That is, all edges are directed starting from the prerequisite math course node and ending at the other node.

In particular, we will take a close look at part of the curriculum at Georgia Southern University. First recall that, Georgia Southern University's Mathematics department offers many courses that are not applicable to the Bachelor of Science degree for Mathematics majors, but are prerequisites of those that are. Such courses include, for instance College Algebra and Trigonometry. In the corresponding evolution network there will be directed edges from College Algebra to Trigonometry and Calculus I. The structural properties of this network will be examined. With the understanding of the network, we propose ways to develop alternatives to the current assessment approaches.

1.5 Summary

In chapter 2 we will start with a brief survey of evolutionary networks and their applications. Most of the content there, although not directly related to our study, shed some light on how evolutionary networks are applied to achieve various goals.

In Chapter 3 we provide a detailed analysis of the mathematics curriculum, based on which the “evolutionary network” is constructed and presented.

In Chapter 4 we briefly discuss how one may employ various concepts in phylogenetic analysis in our network and how to apply them to potential curriculum design.

CHAPTER 2
EVOLUTION NETWORKS

2.1 Overview

Networks of evolution arise in a wide variety of applications such as the World Wide Web, social networks, and communication networks. Generally speaking networks are simply (directed) graphs that often have weights associated with edges.

The recent popularity of dynamic social networks has led to a significant interest in the analysis of evolving networks [2]. Various approaches and techniques in such analysis has several applications, such as, trend analysis in social networks [25, 32, 56, 57, 6] and dynamic link prediction [1, 54, 44, 45]. Networks from practical applications usually evolve in a wide variety of ways that lead to different kinds of evolution semantics.

Analysis of evolving networks can be generally divided into two distinct categories that are not necessarily disjoint:

(a) Maintenance Methods; and

(b) Analytical Evolution Analysis.

The community detection problem is an example, which falls into both categories.

Not all networks evolve at the same rate:

- In email networks, transient links are added to the network in a matter of seconds (corresponding to emails between participant nodes).
- In bibliographic networks, edges are added to the network every few weeks or even months.

For this reason, the evolution of these networks require different kinds of analysis:

- Slowly Evolving Networks: These are networks that evolve slowly over time. Consequently the so called snapshot analysis can be used very effectively.

- Streaming Networks: These are networks that are created by transient interactions such as communication networks. These networks, when represented as graph streams, typically require real-time analytical methods. Such scenarios could arise in the context of streams of objects, edges [58], or linked data streams [31].

2.2 MODELS

In this section we briefly outline few different time-evolving models. Numerous such models exist to serve different kinds of graph analysis problems such as clustering, classification, influence analysis, and link prediction.

Clustering and Community Detection: One of the earliest methods for evolutionary clustering was proposed in 2006. It balances two important objectives while performing the online clustering process:

- the newly formed data clusters should accurately reflect the data at the current time step, and

- the clusters formed at current time step should be closely related to the clusters formed at previous time step.

The former is known as *consistency* and the latter is *smoothness*.

Low Rank Approximation: This model is crucial for identifying the community structures and anomalies in networks. Like other models, there are different kinds of low rank approximations such as Singular Value Decomposition (SVD), matrix factorization, CUR decomposition, Compact Matrix Decomposition (CMD).

Classification: Another type of problem, called node classification, is when the labels of some of the nodes in an evolving network are known and used to predict the labels of the other nodes. The work in [3] proposes a dynamic method for classification of content-based networks. There, a random-walk based approach is used in which the probability of a node being visited is used to determine the class label.

Link Prediction: Link prediction is one of the fundamental problems in the analysis of networks. The link prediction problem simply attempts to determine the links that are most likely to be added to the network in the future. For instance, the Web and social networks are continuously evolving over time, with new nodes and links being added or removed over time.

Tensor Factorization: Tensors are higher order extensions of matrices, data cubes, or multidimensional arrays. Typically, a time evolving graph is simply a tensor whose first two dimensions represent an adjacency matrix and the third dimension captures the sequence of such adjacency matrices representing the evolution of the network. Tensor factorization is then used in link prediction problems.

2.3 Applications

Before applying evolutionary network to our study, in this section we first survey some of the most well-known applications in this section. We will focus on how the models are designed rather than the specific details of the methodology for each application. An overview of the key applications can be found in Table 2.1. For references on the discussion of different kinds of evolution analysis in the context of different kinds of networks, we suggest [10] and [24].

World Wide Web

The Web graphs are approximate topology of the Web created by search engines. Such approximate structures are snapshots used to answer search engine queries. That is, when keywords that are put in, "Page Rank" are computed from the structure of the Web graph.

Domain	Work
World Wide Web	[21, 22, 24, 40]
Telecommunication Networks	[8, 9, 23, 36]
Communication Networks	[21, 29]
Road Networks	[18, 38]
Recommendations	[5, 7, 29, 32, 42, 44, 54]
Social Network Events	[6, 16, 35, 43, 46, 47, 53, 59]
Blog Evolution	[25, 34, 37, 39]
Computer Systems	[12, 30]
News Networks	[33, 57]
Bibliographic Networks	[15, 20, 28, 50]
Biological Networks	[14, 17, 24, 48, 49, 52, 55]

Table 2.1: List of Key Applications of Evolutionary Network Analysis

It is not difficult to imagine that such a graphical structure continues to evolve over time. A study of the evolution of different Web ecologies with the use of visual analysis is provided in [22].

In [40] the authors proposed five similarity schemes for measuring the similarities among the different graph snapshots, three of which are adapted from existing similarity measures [19].

Telecommunication and Mobile Networks

Telecommunication and Mobile Networks can be easily modeled by graphs, where the nodes correspond to participants of particular communications and the edges represent the interactions between participants or simply network connections. In [36], important features of the mobile phone graphs were identified, based on which different ways to model these features were provided.

The evolution of the interactions in mobile networks often corresponds to important network events. In [8] the authors observed that important structural properties of the network could have strong influences on the tie strengths and link persistence between individuals. In [9], an algorithm that operates on a time-varying network of agents was proposed. The algorithm is designed to determine anomalous points in time when agents significantly change their behavior. The problem of finding surprising patterns in the call duration of mobile phone users is also discussed in [23].

Communication Networks

The study of communication networks is very similar to the previous topic. When a fault occurs in a communications network, it usually introduces changes in the routing topology of the network. As a simple example, when an IP router fails, all paths that pass through this particular router will be rendered useless. Such networks has to “evolve” in order to adapt to the occurrence of faults.

The work in [21] uses the regions of correlated spatiotemporal change to identify the root cause of communication network faults. Likewise, the detection of repeated links between nodes has also been used [29].

Road Networks

Evolving network analysis is also important for studying road networks. For instance, [38] models the set of roads as a network and the traffic on the roads as edge weights. Then the evolution is measured in terms of the edge weights while network maintains the same structure.

A related problem is that of mining *heavy subgraphs* in evolving networks [18]. This is somewhat related to the well-known minimal spanning tree problem. These heavy subgraphs are regions in the network where the edge weights are generally very high. It was shown that finding the heaviest dynamic subgraph is NP-hard [18]. An algorithm MEDEN is proposed in [18].

Social Networks

There are mainly two types of problems in the study of Social Networks:

- Social network recommendation:

  The problem of link prediction [45, 51, 11, 41] is directly applied in social networks to suggest friends for different users. Such technologies can also be used for product recommendations in social media networks.

- Social network event detection:

  The “events” of interest include unusual tweets, meetings, or changes in trends in the content of the underlying network. The question of how to detect events was examined in [6].

Other Applications

Lastly we briefly mention some other applications:

- Computer systems:

  A method for (evolutionary) anomaly detection in computer systems with the use of analytical modeling is discussed in [30].

- Blog evolution:

  In this problem each blog post is a node, and a hyperlink between two blog posts can be considered an edge.

- News networks:

  News networks [57] are, in fact, very similar to blog networks. A significance-driven framework was proposed in [57].

- Biological networks:

  Last but not least, in protein-protein interaction networks [55] and [48], nodes correspond to proteins and edges correspond to interactions between them. The evolution of such networks has been one of the most studied questions in recent years.

CHAPTER 3

BUILDING EVOLUTIONARY NETWORK OF MATH CURRICULUM

In this chapter we explore the interconnections (and hence the network) of the curriculum of an undergraduate student. That is, we consider courses that one has to successfully go through before earning a degree in a certain major. The curriculum of any Science, Technology, Engineering and Mathematics (STEM) majors are very structured. Hence, there is a strong relationship between the course sequence that one has to follow in order to complete the degree.

In particular, mathematics is a very well established and historic major. So the 'evolution' of specific courses is rare, except for the ones that are computer based. On another note, the courses in mathematics are very well established, at least in the undergraduate level, which is our current focus. We are interested in the network structure between courses. Such a structure outlines the possible paths a student could be taking in order to complete his or her degree, and such structures changes as the student or the academic advisors continue to make choices along the paths.

3.1 General information

A typical course sequence for a mathematics major is uniquely oriented. As a result, the links are unidirectional. Therefore, the network is similar in nature tp social network event detection, or blog evolutions, or news networks.

It is, however, not as dynamic as those three examples, and is easily predictable. From a student's perspective, the predictability depends on the success or failure of a student in a prerequisite class.

Also, our potential network is not comparable to any rapidly evolving network like modern day's tweets storm, follow-up blogs or facebook post in social media. The time scale in our case is discrete (by semesters). Most of the universities has policies of not accepting any class taken a certain number of years ago. This limits the size of substructures in our network that we can allow, a student to follow.

In this chapter we consider a typical mathematics undergraduate curriculum with time (and financial) limitations. The students are expected to graduate in four years, eight semesters (excluding summer). The connection between the courses is obvious. One has to be proficient (grade ”C” or better) in the prerequisite courses in order to be successful in advanced level courses.

In our network, it is expected that the classes are actually the links, and the nodes are essentially the real status of the student. In reality we could not designate the statuses freshmen, sophomore, junior or senior according to the specific courses they have taken. On the contrary, those designations are reserved for number of credits taken rather than the real standing of the student in actual curriculum.

Currently, these designations are not quite consistent with the student's progression in the actual program. In this current research, therefore, we are NOT going to use these common designations. Instead, we suggest four different stages of progression based on the student's background and they are:

- Pre-beginner (New);
- Beginner;
- Intermediate; and
- Advanced.

This assignment of status will truly reflect the progression of the student within the degree program. In that respect, this network is analogous to the Road Network.

3.2 Network of the mathematics curriculum

In this section, we lay out a mathematics curriculum map as a network.

Recalling earlier description, we note that the network analysis in case of course curriculum is considered “static”. The changes do occur, but they are slow enough, so that the assumption of the “static” network can be justified.

As a case study, we consider the current curriculum of the Bachelor of Science in mathematics degree program at Georgia Southern University. There are several courses that mathematics department offers that do not apply to the degree, such as developmental mathematics courses (below 1000 level), and foundation courses: College Algebra (MATH 1111), Trigonometry (MATH 1112) and Pre-calculus (MATH 1113).

The Mathematics course sequence chart is presented here as network of courses in Figure 3.1 and Figure 3.2.

The link between the two classes is established based on their relations. That is, if a course A is a prerequisite of the course B then there is a link from A to B. These links are established following the curriculum description found in undergraduate catalog. All these documents are published in the university register’s office website

(catalog.georgiasouthern.edu/undergraduate/science-mathematics/mathematical-sciences).

At the bottom of the network lies the developmental courses, which are invisible to general students. The starting point may vary depending on the background skill of the student entering the program. Also, some of the upper-level courses are offered for mathematics education concentration, which are not shown here for simplicity.

Figure 3.1 mainly shows the prerequisite structure up to the elective courses and covers all required courses. It also includes the courses that will not apply to the degree but needs to be satisfied in order to start from Calculus I. Figure 3.2 complements Figure 3.1, and it shows the prerequisite structures for the elective courses. One can combine these two charts into one. However, for clarity two separate charts have been presented.

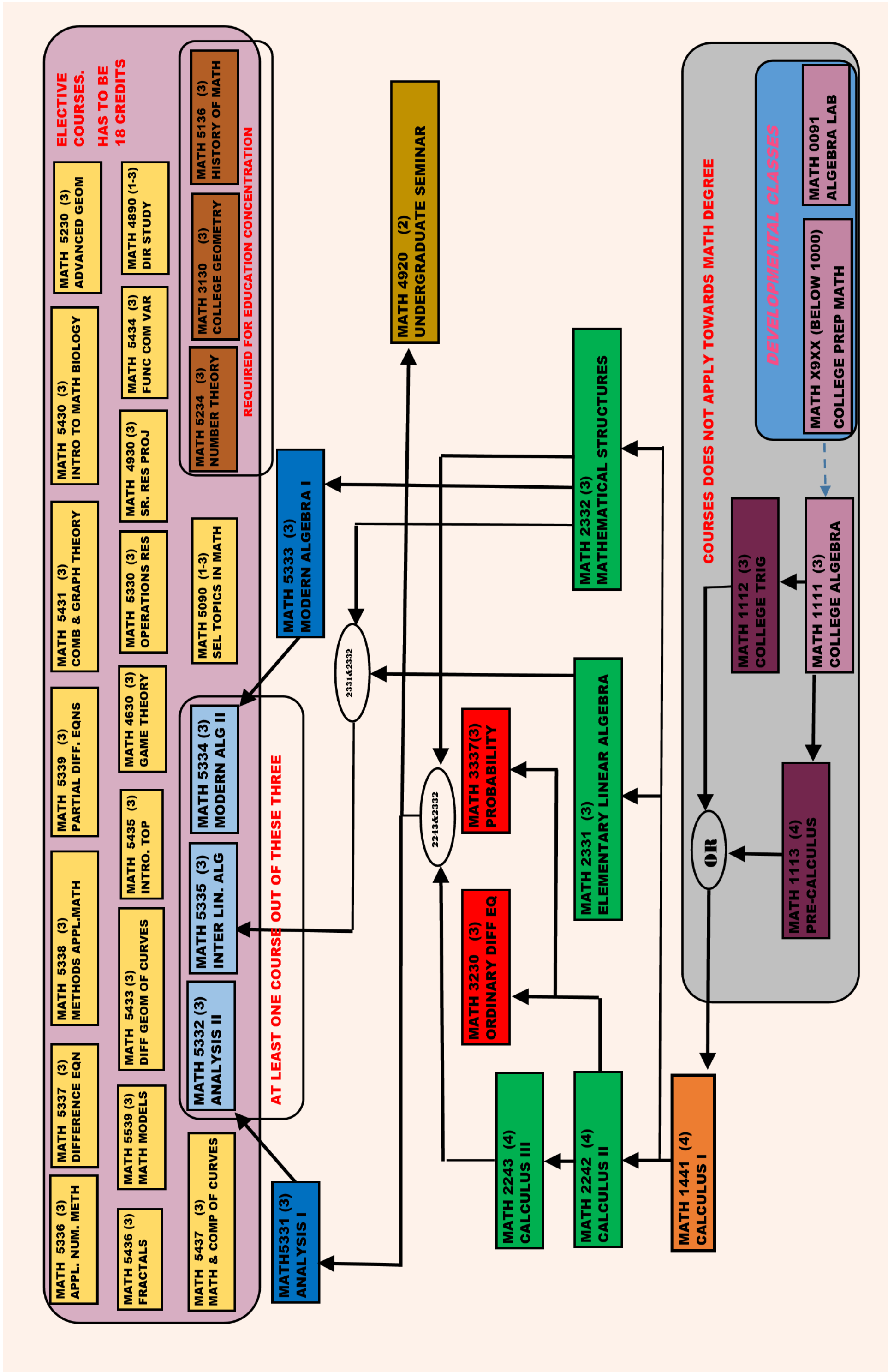

Figure 3.1: Math curriculum structure showing prerequisites until electives.

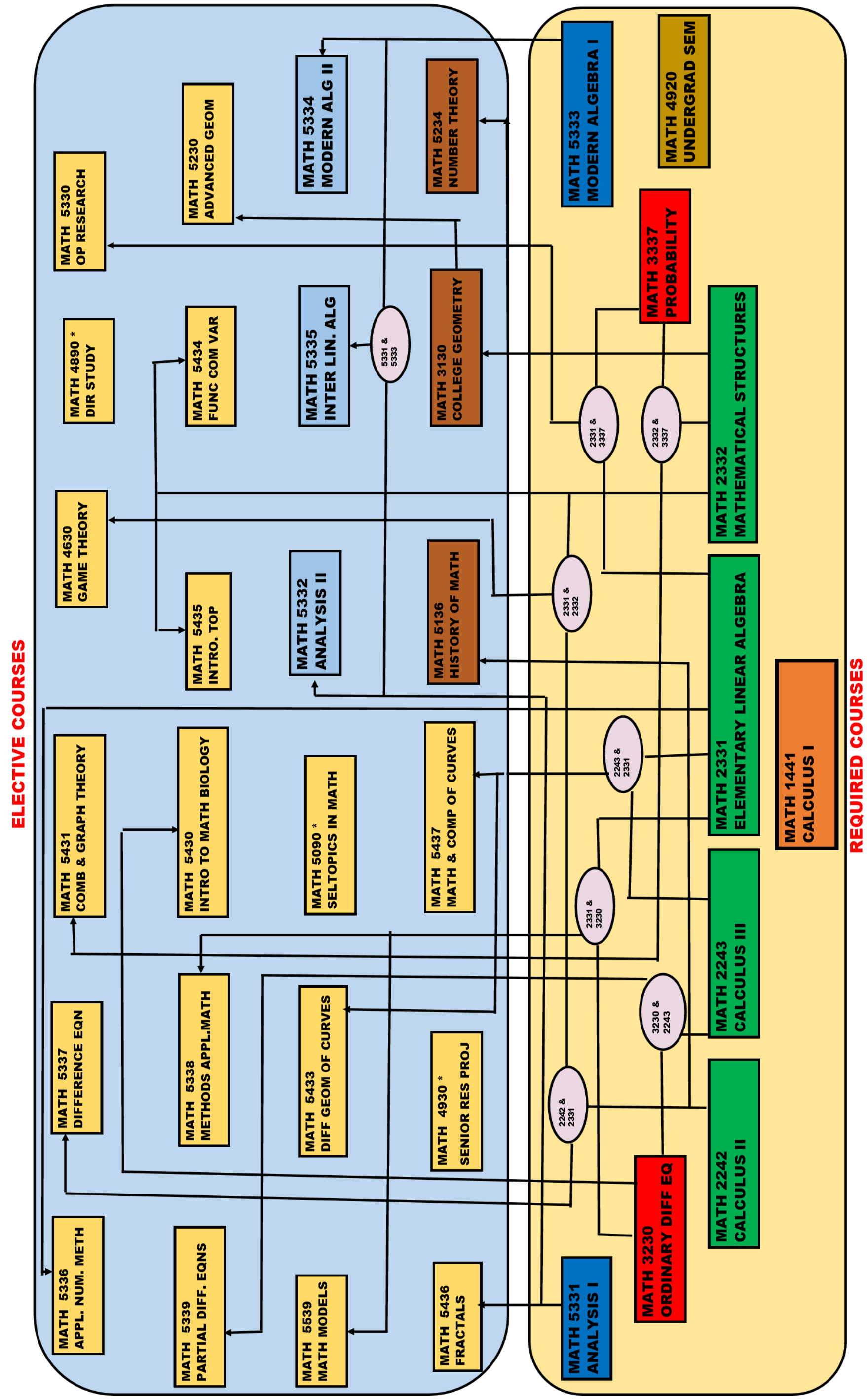

Figure 3.2: Math curriculum structure showing prerequisites only for electives

3.3 Description of the mathematics curriculum network

The above network structures show the location of different mathematics courses within the Mathematics curriculum network. Each block represents one course (with the exception of developmental courses, below 1000 that are grouped together). Courses are represented with course numbers and course names. Also the number within the parenthesis indicates the number of credits. The arrowhead pointed to a block indicates the prerequisite of that specific course. For example, the arrow originating from Calculus I and ending at Calculus II indicates that Calculus I is a prerequisite of Calculus II. If there is more than one prerequisite, then the prerequisites are joined by a circle (or ellipse) with the course number. For example, MATH 2243 and Math 2332 are prerequisites for Math 4920 (Undergraduate Seminar).

Two of the courses are of variable credits: Selected Topics in Math (MATH 5090) and Directed Study (MATH 4890). Also note, these two do not have any course prerequisite as such. They all require permission of instructor. Similarly, Senior Research Project (MATH 4930) does not have any course prerequisite but its prerequisite is completion of 15 credit hours of upper level mathematics, statistics and/or computer science.

Similarly, if there is a choice of prerequisites, then they are joined by an ”OR” circular block. For example, one has to complete MATH 1112 (Trigonometry) or MATH 1113 (Pre-calculus)in order to qualify for Calculus I (Figure 3.1) .

In Figure 3.1, the courses that do not apply to the degree but are required to be satisfied are grouped together at the right hand bottom corner. All the elective courses are grouped together as well. They are situated at the upper part of Figure 3.1 and Figure 3.2. Also there are some special grouping within the elective courses as indicated in Figure 3.1. They can be viewed as subsets of the set of elective courses. For example, there are three elective courses (MATH 5332, MATH 5335 and MATH 5334) from which one has to choose at least one. These three courses are grouped together. There are also three courses in the elective section (MATH 5234, MATH 3130 and MATH 5136), where a math major with education specialization has to have. The first subset colored in blue and the second subset colored in brown. All other elective courses are in yellow box.

3.4 STUDENT STAGES IN THE NETWORK

As mentioned at the beginning of this chapter, there are three distinct landmarks within the mathematics program:

- ”Beginners” as they are ready to start Calculus I;
- ”Intermediate” as they have completed four beginner-level courses (Calculus I, Calculus II, Elementary Linear Algebra and Math Structure), and are ready to take on other required courses for the program;
- ‘Advanced’ when they have completed all the required courses and ready to take on elective courses.

In addition, it can be noted that there are five different stages of a student who just got admitted in the college has to go through in order to graduate with a mathematics degree. First of all, we make a distinction here between the college beginners and the math beginners. College beginners are those who just got admitted to the college and math beginners are those who are ready to take Calculus I. Hence, the first stage at the math program is the math beginners (not necessarily college freshmen). As can be seen from Figure 3.1, one has to ultimately satisfy either a combination of college algebra and trigonometry, or pre-calculus to become a Calculus I ”ready”. The math program begins right at that point.

Referring to the same figure, it can be noted that Calculus I is the very first course in the sequence. Calculus II, Elementary Linear Algebra and Mathematical Structures all have only one prerequisite, which is Calculus I. Hence, a student can take Calculus I and immediately followed by these three courses and it will take a student two semesters, one academic year. Following completion of this course sequence a student reaches next level in the mathematics program, which is intermediate stage. From that point, till a student completes all the required courses, he/she remains in the intermediate state. Once they complete four courses (Calculus I, Calculus II, Linear Algebra and Math Structures, fourteen credits in total) as beginners, four other required courses are accessible to them. (To access to other two required courses one has to complete Calculus III). Once they complete all the required named courses, they reach the advanced state within the program. At that point ONLY the elective courses have to be taken to complete the degree. While they are taking the elective courses, they remain in the Advanced state within the program. Once they complete their elective courses, they are ready to graduate. which is the final landmark. With these stages in mind (pre-beginner, beginner, intermediate and advanced) a generic tree for mathematics program has been shown in Figure 3.3.

3.5 A BREAK-DOWN OF THE CURRICULUM REQUIREMENTS

3.5.1 PRE-BEGINNER

It can be noted that the number of credits for the beginners level courses is fourteen. There are six intermediate-level courses totaling eighteen credits. Similarly the total number of credits for advanced-level courses is eighteen and it can be accomplished by taking six courses (unless someone take a combination Special Topics and Directed Study to make up for three credits). Considering this course structure within the Mathematics program three other figures are created to reveal the details at each stage. Figure 3.4 shows the courses at the pre-beginner (pre-Math) stage and how it leads to the Calculus I, the very first course in the beginner stage.

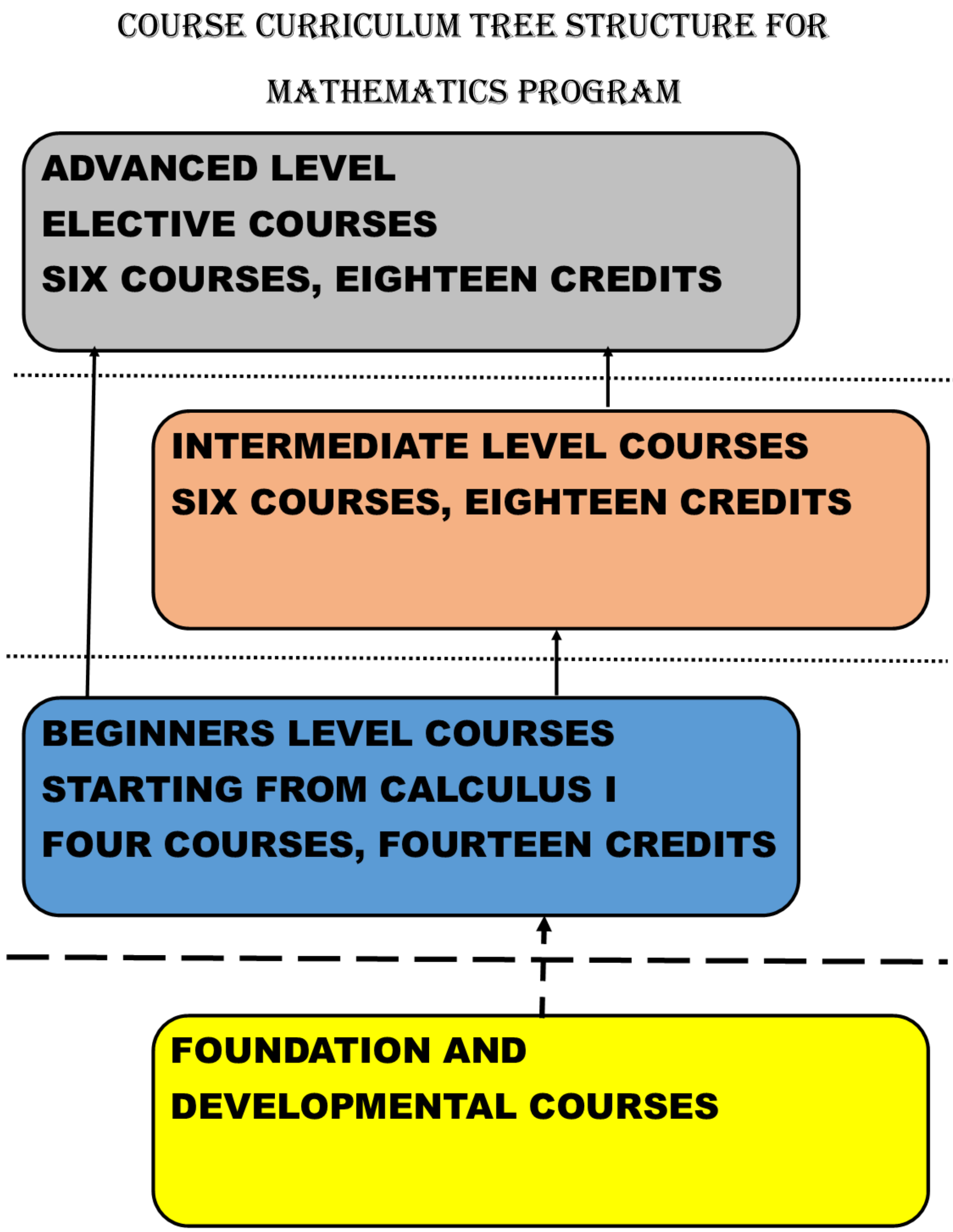

Figure 3.3: Math course curriculum tree structure using blocks.

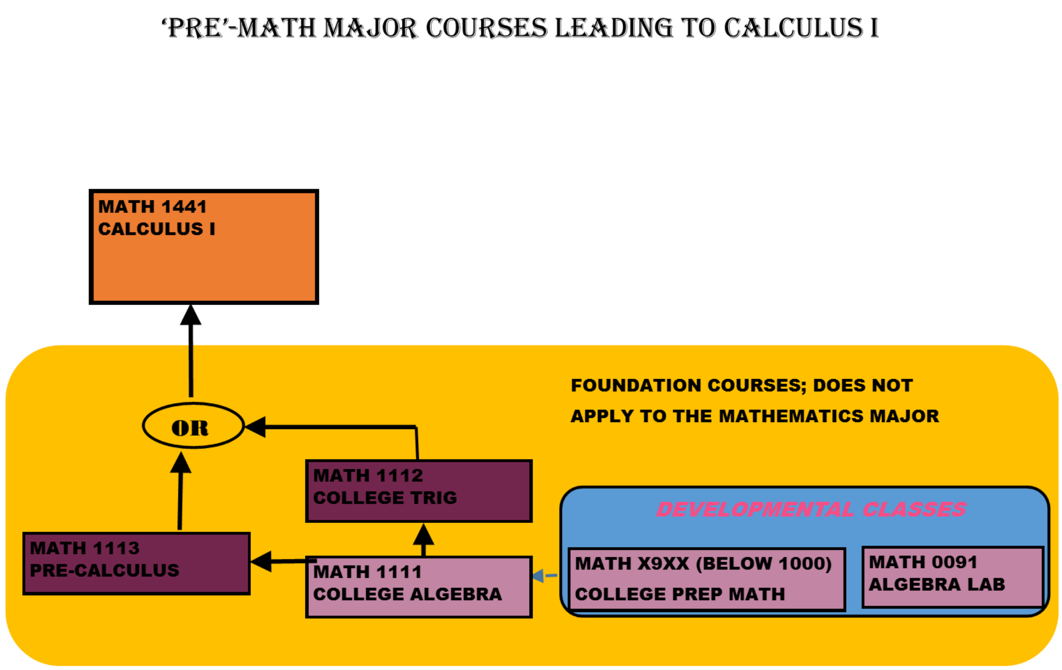

Figure 3.4: Structure for pre-beginner courses leading to Calculus I.

3.5.2 Beginner and intermediate levels

The next diagram, Figure 3.5, shows links between the beginners courses and the intermediate-level courses.

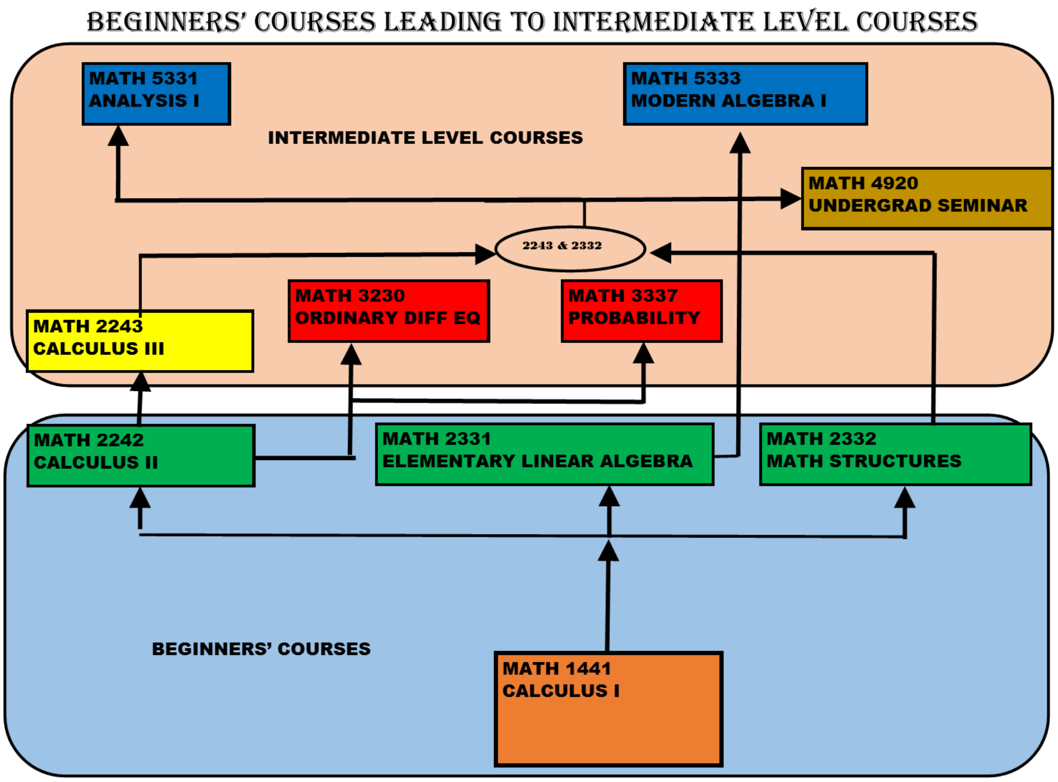

Figure 3.5: Beginners' courses leading to intermediate level courses.

3.5.3 Advanced level

Finally, the Figure 3.6 shows the advanced level courses and their link with both beginner and intermediate-level courses.

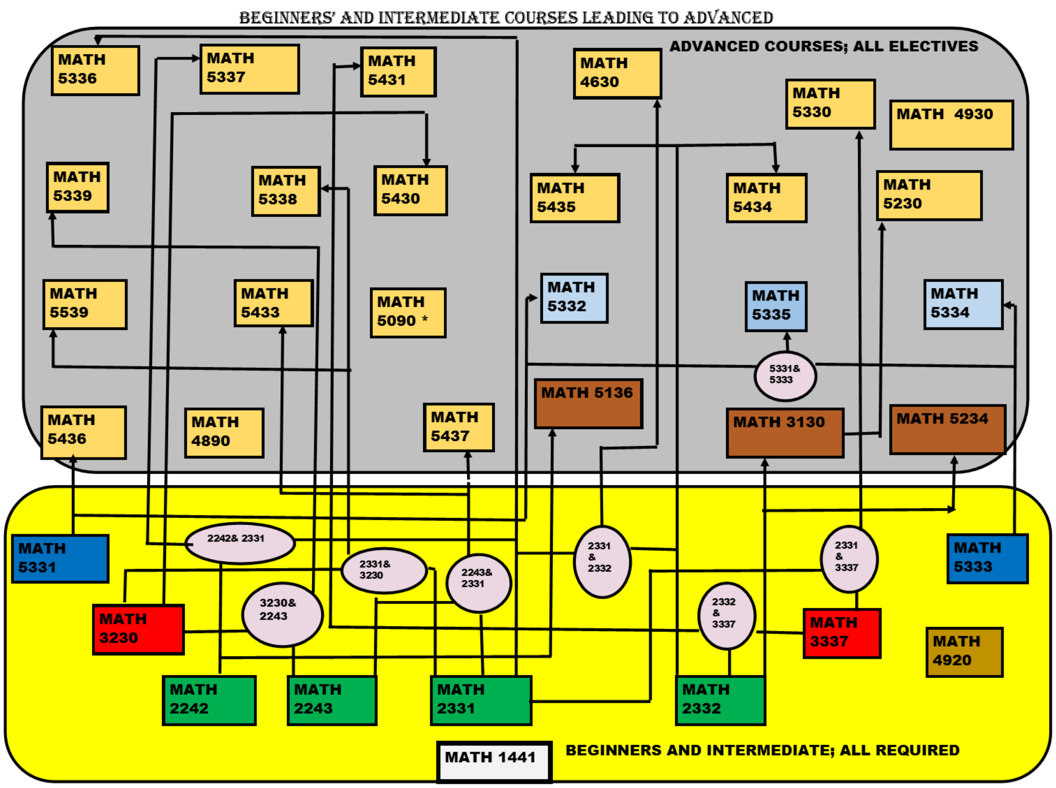

Figure 3.6: Beginners and Intermediate courses leading to advanced level courses.

3.6 IDENTIFICATION OF KEY COURSES

We are exploring some of the features of the tree structure of mathematics course sequence. As described earlier, the beginning level courses are very linear. However, when it comes to the intermediate level, there are several choices. Some of the courses are prerequisites of more required and elective courses than the others. In this section we identify those courses as key courses. Table 3.1 shows the list of courses that serve as prerequisites for three or more higher level courses. Except for two, all other required courses serve as prerequisites for at least two others courses in the curriculum and they are not listed in the table. Here "number of PRC" denotes the "number of courses served as Prerequisites for Required Courses" and "number of PEC" denotes the "number of courses served as Prerequisites for Elective Courses". The table also shows the break up between the required and elective courses. For example, Calculus II is prerequisite of three other required courses and two other elective courses. Elementary Linear Algebra serves as prerequisite for no other required course, however, it is prerequisite for nine elective courses. Similarly, Mathematical Structure is prerequisite for three other required courses and seven other electives courses. Among intermediate level courses, Calculus III serves as prerequisite for two other required and three other elective courses. Ordinary Differential Equation serves as a prerequisite for four other elective courses.

It can be noted that, out of all six intermediate level courses, Ordinary Differential Equation and Calculus III feature in the table 3.1. Hence, strategically it is advisable to give priority to those two courses over the other four courses at the intermediate level.

Courses at the beginners level	Number of PRC	Number of PEC	Total
Calculus I (Math 1441)	3	0	3
Calculus II (Math 2242)	3	2	5
Elementary Linear Algebra (Math 2331)	0	9	9
Mathematical Structures(Math 2332)	3	7	10

Courses at the intermediate level	Number of PRC	Number of PEC	Total
Calculus III (Math 2243)	2	3	5
Ordinary Diff Equation (Math 3230)	0	4	4

Table 3.1: Key courses

CHAPTER 4

POTENTIAL APPLICATIONS OF THE CURRICULUM NETWORK

The Mathematics course curriculum network, as described in Chapter 3, has several applications. In this chapter three of the major applications will be discussed. They are:

- Academic advisement;
- Resource allocation;
- Assessment for Mathematics Skill Level.

The first one is obviously important from a students perspective aspiring at a Mathematics degree. The second one has lot of significance from an administrators point of view regarding differential resource allocation for various courses. And the third one would be significant from the programs assessment standpoint.

We will also briefly discuss some previously surveyed topics and how they can be applied to the curriculum networks.

4.1 Academic Advisement

Students coming into the college considering Mathematics as their degree program need a guided advisement for the course sequence in order to complete their degree programs in a time-efficient way. Based on their current Mathematics level skill, they can be assigned to any of the foundation level classes (college algebra, pre-calculus, trigonometry etc) or even remedial developmental classes or they can be placed directly into Calculus I or even Calculus II (if they qualify through their high school curriculum). From that point, until they graduate, it is important that they complete the courses in a highly sequential manner. It is strongly recommended that they complete the basic level courses before they move on to their intermediate level. The same argument can be applied in transitioning from intermediate level courses to advanced level courses. Similarly, students specializing in Mathematics Education or Statistics can be guided accordingly. In previous figures (Figure 3.1, Figure 3.2, Figure 3.5), Mathematics Education advanced level required courses are indicated and they are highlighted with brown color. Other Mathematics Education electives specialization courses are sensitive to the job requirement, high-school, middle-school etc. Many of the statistics specialization courses start with STAT numbers and out of scope for this current work, although similar analysis to the one presented here can be performed for Statistics courses.

4.2 RESOURCE ALLOCATION

In almost all universities, Mathematics departments provide service courses for other majors. So it is always a challenge to allocate resources for various courses and it is often heavily dependent on the population of the other majors. For example, most of the engineering majors require calculus sequence, while many of the health science related majors require basic Statistics course. However, if we narrow down and only look from the Mathematics major perspective, there are several courses which are in higher demand in comparison to others. The Table 3.1 shows that Mathematical Structure is prerequisite for ten other courses and Elementary Linear Algebra is prerequisite for nine other courses including electives courses. Those two courses can easily be a ”bottleneck” for graduation. With that in mind, it is highly recommended that one section of these two courses needs to be offered only for the Mathematics majors students, depending on the number of students majoring in Mathematics. On the other hand, the Undergraduate Seminar is not a prerequisite for any course but it is required for graduation. So a student can have at least a few semesters to take that course while taking other courses to complete the degree.

4.3 Assessment for mathematics skill level

Based on the structure developed in the previous chapter, we proposed four mathematics skill levels:

- pre-beginner;
- beginner;
- intermediate; and
- advanced.

Just as general assessment is a big part of the quality enhancement plan and the development of the university, assessment of mathematical maturity will greatly help us with the placement of students and allocation of resources.

Mathematics assessment test can be conducted at each of the above listed transition gates:

- pre-beginner to Calculus I;
- beginner to intermediate; and
- intermediate to advanced elective level courses.

These assessment data will be much more useful than the data collected from a single course or for that matter, assigning the students Freshmen, Sophomore, Junior, Senior levels based solely on their college credits.

4.4 Application of spanning tree and sequence alignment

Earlier in this thesis, among many other useful concepts and approaches are the spanning trees and sequence alignment. In this section we will briefly explore how they, and other related mathematical tools, can be used in our application to the curriculum network.

4.4.1 Spanning trees

As mentioned before, the main problem related to spanning trees deals with finding the spanning tree with minimum total weight. Such weights may be assigned to edges connecting a pair of classes depending on how important and urgent it is, and finding a minimum spanning tree under such conditions will give us an efficient way of covering all classes. This is useful when the prerequisites are under exclusive "or" conditions.

Alternatively, one may assume that a student comes into the program (or at certain point of their academic career) with some credits already earned. We now seek a sub-structure (not necessarily a spanning tree), in the curriculum network, that contains this given set of nodes/courses, and another set of "target courses" (that is, the ones required for the particular degree or concentration). Such sub-structures are called *Steiner trees* when they are acyclic.

4.4.2 Sequence alignment

As we have seen earlier, sequence alignment is essentially the comparison between two sequences of entries chosen from a given alphabet. In our case, the set of alphabets is the collection of available courses.

Suppose, now, that two students with different backgrounds come into the program at the same time. We will represent each of them with the set of "fundamental courses" (i.e. the courses that serve as prerequisites for others in the curriculum), ordered from the most elementary to the most advanced. If a course appears in one student's transcript but not the other, we will save the spot and leave an "empty" entry there.

Thus to compare two students' background we compare the two ordered sequences. How similar two students are (and whether they can share a similar academic plan in the future) depends on how "aligned" these two sequences are.

On the other hand, one may simply construct a standard sequence (corresponding to the ideal collection of mastered courses) at each phase and compare each student's background with those to determine their placement.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE WORK

In this thesis we considered the concept of phyloenetic trees and the more generalized version, the evolutionary networks. We started with a survey on known problems and results on these concepts, related mathematical tools, and their applications.

We then attempted to apply this concept to the mathematics curriculum. To do this, we first examined the relationship between various courses and constructed various diagrams to present these relationships. These diagrams, when combined, gives us an interesting evolutionary network to work with.

Based on the network constructed we point out several potential applications, including the use of spanning trees, sequence alignments, and related tools.

The theoretical ideas proposed here appear to be sound and the applications presented seem to be promising. The natural next step is, of course, to try to implement this system in the decision making of academic advisement, placement of students, resource allocations, etc.

REFERENCES

[1] E. Acar, D. Dunlavy,and T. Kolda, 2009. Link prediction on evolving data using matrix and tensor factorizations. ICDMW, IEEE, 262–269.

[2] C. Aggarwal, On classification of graph streams, *SDM* (2011) 652–663.

[3] C. Aggarwal and N. Li 2011. On node classification in dynamic content-based networks.SDM, 355–366.

[4] C. Aggarwal, Y. Li, P. Yu, and R. Jin 2010. On dense pattern miming in graph streams. PVLDB 3, 1(2010), 975–984.

[5] C. Aggarwal, S. Lin, and P. Yu, 2012a. On influential node discovery in dynamic social networks. SDM, 636–647.

[6] C. Aggarwal and K. Subbian 2012. Event detection in social streams.SDM, 624–635.

[7] C. Aggarwal, Y. Xie and P. Yu 2012b. On dynamic link inference in heterogeneous networks.SDM, 415–426.

[8] L. Akoglu and B. Dalvi 2010. Structure, tie persistence and event detection in large phone and SMS networks. MLG Workshop.ACM 10–17.

[9] L. Akoglu and C. Faloutsos 2010. Event detection in time series of mobile communication graphs. Army Science Conference.

[10] L. Akoglu and C. Faloutsos 2013. Anomaly, event, and fraud detection in large network datasets. WSDM, 773–774.

[11] M.A. Hasan, V. Chaoji, S. Salem, and M. Zaki, 2006. Link prediction using supervised learning. SDM Workshop on Link Analysis, Counter-terrorism and Security.

[12] R. Albert and A. Barabasi, 2000. Topology of evolving networks: Local events and universality. Physical Review Letters 85, 24(2000), 5234–5237.

[13] R. Albert and A. Barabasi, 2002. Statistical mechanics of complex networks. Reviews of Modern Physics 74, 1(2002), 47.

[14] S. Asur, S. Parthasarathy, and D. Ucar, 2007.An event-based framework for characteririzing the evolutionary behavior of interaction graphs. KDD. ACM, 913–921.

[15] A. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A.Schubert, and T. Vicsek, 2002. Evolution of the social network of scientific collaborations. Physica A 311, 3(2002), 590–614.

[16] A. Beutel, W. Xu, V. Guruswami, C. Palow and C. Faloutsos, 2013. Copy-Catch: Stopping group attacks by spotting lockstep behavior in social networks. WWW, 119–130.

[17] A. Beyer, P. Thomason, X. Li, J. Scott and J. Fisher, 2010. Mechanistic insights into metabolic disturbance during type-2 diabetes and obesity using qualitative networks. In Transactions on Computational Systems Biology XII, Springer, 146–162.

[18] P. Bogdanov, M. Mongiovi, and A.K.Singh, 2011. Mining heavy subgraphs in time-evolving networks. ICDM, IEEE, 81–90.

[19] H. Bunke, P. Dickinson, M. Kraetzl, and W. Wallis, 2006. A Graph-Theoretic Approach to Enterprise Network Dynamics. Vol 24, Birkhauser, Boston.

[20] D. Chakrabarti, R. Kumar, and A. Tomkins, 2006. Evolutionary clustering. KDD.554-560.

[21] J. Chan, J. Bailey, and C. Leckie, 2008. Discovering correlated spatio-temporal changes in evolving graphs.Knowledge and Information Systems 16, 1 (2008), 53–96.

[22] E. Chi, J. Pitkow, J. Mackinlay, P. Pirollo, R. Gossweiler, and S. Card, 1998. Visualizing the evolution of web ecologies. ACM SIGCHI Conference, 400–407.

[23] P. Vaz De Melo, L. Akoglu, C. Faloutsos, and A. Loureiro, 2010. Surprising patterns for the call duration distribution of mobile phone users. In Machine Learning and Knowledge Discovery in Databases. Springer, 354–369.

[24] S. Dorogovtsev, J. Mendes, 2003. Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press.

[25] M. Goetz, J. Leskovec, M. McGlohon, and C. Faloutsos, 2009. Modeling Blog Dynamics, ICWSM.

[26] I. Gunes, Z. Cataltepe, and S. Gunduz-Oguducu, 2013. GA-TVRC-Het: Genetic algorithm enhanced time varying relational classifier for evolving heterogeneous networks. DMKD. 1-32.

[27] M. Gupta, C. Aggarwal, and J. Han, 2011. Finding top-k shortest path distance changes in an evolutionary network. SSTD. 130–148.

[28] M. Gupta, C. Aggarwal, J. Han, and Y. Sun, 2011b. Evolutionary clustering and analysis of bibliographic networks. ASONAM, 63–70.

[29] Z. Huang and D. Lin, 2009. The time-series link prediction problem with applications in communication surveillance. INFORMS Journal on Computing 21, 2(2009), 286–303.

[30] T. Ide and H. Kashima, 2004. Eigenspace-based anomaly detection in computer systems. KDD, 440-449.

[31] D. Le-Phuoc, J.X. Parreira, and M. Hauswirth, 2012. Linked stream data processing. In Reasoning Web: Semantic Technologies for Advanced Query Answering. Springer, 245–289.

[32] J. Leskovec, L.A. Adamic, and B.A. Huberman, 2007.The dynamics of viral marketing. ACM TWEB 1, 1(2007), 5.

[33] J. Leskovec, L. Backstrom, and J. Kleinberg, 2009. Meme-tracking and the dynamics of the news cycle. KDD. ACM,497–506.

[34] J. Leskovec, M. McGlohon, C. Faloutsos, N.Glance and M.Hurst, 2007.Patterns of cascading behavior in large blog graphs. SDM, 551–556.

[35] C. Lin, B. Zhao, Q. Mei, and J. Han, 2010. PET: A statistical model for popular events tracking in social communities. KDD. ACM, 929–938.

[36] S. Liu, L. Li, C. Faloutsos, and L.M. Ni, 2011. Mobile phone graph evolution: Findings, model and interpretation. ICDM Workshops. 323–330.

[37] M. McGlohon, J. Leskovec, C. Faloutsos, M. Hurst and N. Glance, 2007. Finding patterns in blog shapes and blog evolution. ICWSM.

[38] M. Mongiovi, P. Bogdanov, R. Ranca, E. Papalexakis, C. Faloutsos, and A.K. Singh, 2013. Netspot: Spotting significant anomalous regions on dynamic networks.SDM.SIAM.

[39] H. Ning, W. Xu, Y. Chi, Y. Gong,and T.S. Huang, 2007. Incremental spectral clustering with application to monitoring of evolving blog communities. SDM, 261–272.

[40] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina, 2010. Web graph similarity for anomaly detection. Journal of Internet Services and Applications 1, 1(2010), 19–30.

[41] A. Popescul and L. Ungar, 2003.Statistical relational learning for link prediction.IJCAI Workshop on Learning Statistical Models from Relational Data, Vol.2003. Citeseer.

[42] M.Richardson and P. Domingos,2002. Mining knowledge-sharing sites for viral marketing.KDD.ACM, 61–70.

[43] T. Sakaki, M. Okazaki, and Y. Matsuo, 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. WWW. ACM, 851–860.

[44] P. Sarkar, D. Chakrabarti, and M. Jordon, 2012. Nonparametric link prediction in dynamic networks. ICML.

[45] R. Sarukkai, 2000. Link prediction and path analysis using Markov chains.Computer Networks 33, 1(2000), 377–386.

[46] H. Sayyadi, M. Hurst,and A. Maykov, 2009. Event detection and tracking in social streams. ICWSM.

[47] J. Silva and R. Willett, 2008. Detection of anomalous meetings in a social network. Information Sciences ans Systems. IEEE, 636–641.

[48] R. Sole, R. Pastor-Satorras, E. Smith, and T.B. Kepler, 2002. A model of large-scale proteome evolution. Advances in Complex Systems 5, 1(2002), 43-54.

[49] J.M. Stuart, E. Segal, D. Koller, and S.K. Kim, 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science 302, 5643(2003), 249–255.

[50] Y. Sun, R. Barber,M. Gupta,C. Aggarwal,and J. Han, 2011. Co-author relationship prediction in heterogeneous bibliographic networks.ASONAM. IEEE, 121–128.

[51] B. Taskar, M. Wong, P. Abbeel, and D. Koller, 2003. Link prediction in relational data.NIPS, Vol.15.

[52] S. Teichmann and M. Madan Babu, 2004. Gene regulatory network growth by duplication. Nature genetics 36, 5(2004), 492–496.

[53] H. Tong, Y. Sakurai, T. Rad,and C. Faloutsos, 2008c. Fast mining of complex time-stamped events. CIKM. ACM, 759–768.

[54] T. Tylenda, R. Angelova, and S. Bedathur, 2009. Towards time-aware link prediction in evolving social networks.*Social Network Mining and Analysis Workshop.* 9.

[55] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, 2002. Modeling of protein interaction networks. Complexus 1, 1(2002), 38–44.

[56] C. Wang and L. Chen 2009, Continuous Subgraph pattern search over graph streams. ICDE, 393–404.

[57] L. Yan, J. Wang, J. Han and Y. Wang, 2012, A significance-driven framework for characterizing and finding evolving patterns of news networks. Artificial Intelligence and Computer Intelligence. Springer, 134–141.

[58] Gsketch: On query estimation in graph streams, VLDB 5, 3(2011), 193–204.

[59] Q. Zhao, P. Mitra, and B. Chen, 2007. Temporal and information flow based event detection from social text streams. NCAI, Vol.2, 1501.

[60] *http://doebleygen564s14.weebly.com/how-to-make-a-phylogenetic-tree.html.* (Last accessed 03/31/2018).

[61] *https://www.expii.com/t/minimum-spanning-tree-mst-9822?type=explanation&id=3406.* (Last accessed 03/31/2018).

[62] *https://towardsdatascience.com/pairwise-sequence-alignment-using-biopython-d1a9d0ba861f.* (Last accessed 03/31/2018).

[63] *http://etherealmind.com/network-evolution-hierarchy-needs.* (Last accessed 03/31/2018).