Summer 8-6-2018

MAnanA: A Generalized Heuristic Scoring Approach for Concept Map Analysis as Applied to Cybersecurity Education

Sharon Elizabeth Blake Gatto
sblake@uno.edu

MAnanA: A Generalized Heuristic Scoring Approach for Concept Map Analysis as Applied to Cybersecurity Education

A Thesis

Submitted to Graduate Faculty of the
University Of New Orleans
in partial fulfillment of the
requirements for the degree of

Master of Science
in
Computer Science
Information Assurance

by

S. E. Blake Gatto

B.A. University of South Alabama, 1998

B.S. University of South Alabama, 1998

August, 2018

Acknowledgment

I would like to thank my thesis adviser, Dr. Irfan Ahmed, for his instruction, encouragement, and support. I would like to also thank Dr. N. Adlai A. DePano, and Dr. Minhaz Zibran for serving on my thesis Defense Committee.

Next, extra special thanks to my family, especially to my wonderful, brilliant, patient, and loving sons, Alex and Zach, for not running away from home and putting up with me throughout my graduate studies.

Finally, much appreciation and respect to my university colleagues for the encouragement and inspirations, especially William Johnson, Ryan Joseph, Manish Bhatt, Dr. Brian Roux, Dr. Joe Sylve, and Dr. Golden Richard III.

for Zach & Alex

Contents

List of Figures

List of Tables

Abstract

Concept Maps (CMs) are considered a well-known pedagogy technique in creating curriculum, educating, teaching, and learning. Determining comprehension of concepts result from comparisons of candidate CMs against a master CM, and evaluate "goodness". Past techniques for comparing CMs have revolved around the creation of a subjective rubric. We propose a novel CM scoring scheme called `MAnanA` based on a *Fuzzy Similarity Scaling* (`FSS`) score to vastly remove the subjectivity of the rubrics in the process of grading a CM. We evaluate our framework against a predefined rubric and test it with CM data collected from the *Introduction to Computer Security* course at the University of New Orleans (UNO), and found that the scores obtained via `MAnanA` captured the trend that we observed from the rubric via peak matching. Based on our evaluation, we believe that our framework can be used to objectify CM analysis.

KEY WORDS

Cybersecurity Education, Concept Map Analysis, Fuzzy Set Theory and Logic, Heuristic Scoring, Membership Function

Chapter 1

Introduction

Concept Maps (CMs) are considered a well-known pedagogy technique that allows us to think about the concepts and precisely establish relationships among them [3–5]. They are used in curriculum design, concept organization, evaluation of teaching, teaching strategy design, etc [4]. Structurally, CMs are acyclic graphs which contain nodes characterized by concepts with concepts, and the edges are connections between the nodes which are representative of the relationship between the concepts. It is an important educational tool in a wide variety of fields.

In the cybersecurity education arena, there are a myriad of ways to present the advanced ideas to the student population. Some of these techniques are programmatic [6] while others are based on educational techniques [7–9]. However, not all students are immediately able to comprehend such complex ideas and constructs found in the upper level courses of computer sciences. Rote learning and memorization of ideas may not lead the student to a deeper comprehension of the topic, but perhaps leads only to a superficial understanding, which is not necessarily an acquirement of knowledge. A student would need to be able to acquire an in depth knowledge of the advanced topics in cybersecurity through the development of critical thinking and problem solving skills [6]. CMs have been used as an effective teaching tool since its introduction in 1984 [3, 10] and thus, would be effective aid in cybersecurity education. Comprehension of such concepts require the student to examine the material and evaluate the relationships with skills of critical thinking and problem solving.

In order to provide some clarification to the complicated concepts of cybersecurity education, we investigate the usage of concept maps in the upper level undergraduate and introductory graduate "Introduction to Computer Security" course curriculum at the University of New Orleans (UNO). To better understand how well the students were able to incorporate the knowledge presented, we then analyzed the concept maps both quantitatively and qualitatively. In our research, we find that the "goodness" of a concept map is measured by comparing it with a model concept map created by an instructor or a subject matter expert [11]. However, the comparison mechanisms (discussed in the related work section) were either quite complicated to implement, or were quite subjective via creation of a subjective rubric to score them.

In this work, we propose an analytic framework, called `MAnanA`, to provide a more rigorous and objective assessment to the naturally subjective characteristics of cybersecurity CMs. `MAnanA` works as follows: first it extracts certain content-measure features and structural graph features from CMs, and then it calculates a heuristic metric to evaluate the "goodness" of the concept maps, called the `Fuzzy Similarity Scaling (FSS)`. The `FSS` score is a novel CM scoring metric based on distance measures. It stems from the foundations of *Fuzzy Sets Theory* and *Fuzzy Logic Theory* [12–15], and creates a scoring scheme to better objectify the analysis of concept maps in general. To evaluate the efficacy of `FSS`, we require a medium of establishing ground truth. In order to satisfy this need, we created a rubric to grade the CMs based on specific criteria for the student-submitted CMs in the aforementioned course. Finally, we compared `FSS` to the scores obtained via the rubric and found that the scores obtained via `FSS` conformed quite well to the trendlines with positive peak-matching of the CM scores obtained via the rubric. Therefore, we believe that `FSS` can be used as a better mechanism to compare CMs as well as grade them in lieu of the subjective rubrics.

This thesis is organized as follows: first, it discusses previous works which have been done in the concept maps analysis area. Then, the thesis provides theoretical background and any essential mathematics required to understand the control flow of `MAnanA`, provides details of the data collection and experimentation procedure and presents an evaluation of `MAnanA`, and finally concludes presenting the scope of our future work.

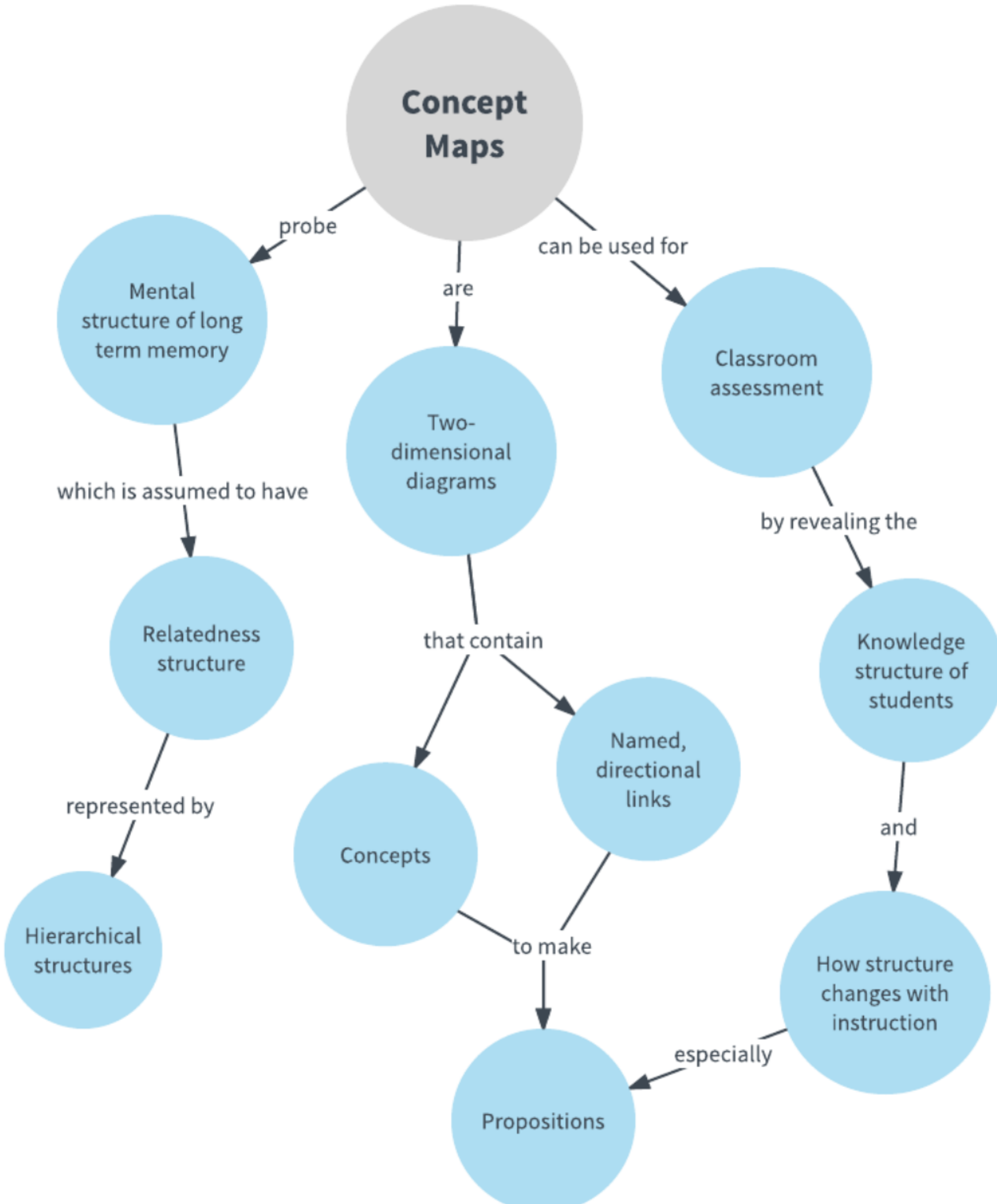

Figure 1.1: Example of a Concept Map [1]

Chapter 2

Related Work

While the use of CMs in the teaching and the learning side of education have been widely researched at different levels, the problem of comparing one CM to another is a problem that has seldom been met in literature. A meaningful way of evaluating CMs is to compare it against the “master CM” provided by the instructor. This process can be very time consuming since automation of this process is hard to achieve. Because of the free form of concepts, relationships, and propositions, detailed grading of concept map elements requires manual work and domain expertise. Scoring of concept maps based on quality of the elements have been studied [16, 17]. They adopted and modified the previous scoring methods to evaluate students’ work. Basically, the instructor created a “master CM”, against which student work were compared to obtain “goodness” of the concept maps and is often assigned by the graders who are familiar with the purpose of the assessment [11]. This method is, however, weak as to the subjective nature of the scoring rubrics although the idea of creating a continuous objective scoring system between *no CM* and the *control CM* is quite interesting.

Moreover, some mathematical approaches have been proposed to compare CMs. Muhling proposes a quality measure based on the intersection of the concept sets [18]. Similar measures have also been proposed by others [19, 20]. Leake also proposes measures used to evaluate usefulness of the concepts as well as contextual similarity [20]. However, since the measures herein proposed are `set` based, a simple rewording of the concepts would render the method as inadequate. In literature, we were unable to find a more robust method that includes *set theory* with other features such that simple ways of skewing results like rewording the concepts would not be enough question their adequacy.

Chapter 3

Proposed Framework - MAnanA

The `MAnanA` framework for analysis and heuristic scoring of the CMs constitutes of *Feature Selection*, *Fuzzy Membership Function Calculation*, and *Fuzzy Similarity Scaling (FSS) Scoring*.

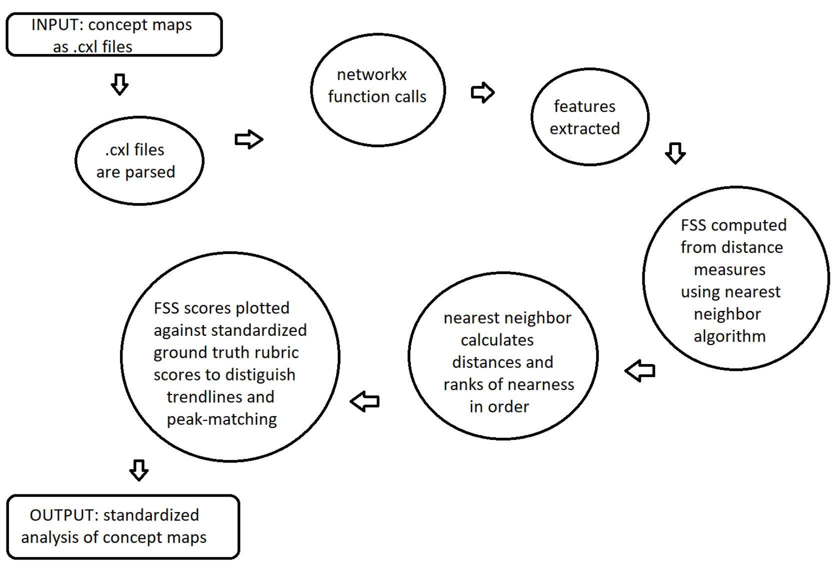

Figure 3.1: `MAnanA` Overview.

3.1 Overview

As seen in Figure 3.1, we examine the processes within the `MAnanA` framework itself. Input of .cxl files to the parser allows the extraction of a particular set of features via the use of Python's sklearn and networkx packages. These features include content measures and graph structure from the CMs as input. Content measure features include number of nodes (concepts), number of edges (linkages), average number of words per node (concept), average number of characters per node (concept), number of crosslinks, number of hierarchies, and highest hierarchy. Graph structural features include distribution of in-out degrees per concept, average of degree centrality, average of betweenness centrality, average of closeness centrality, and cardinality of maximal clique. After feature extraction, their values are used to calculate the score for `FSS` via the nearest neighbor algorithm, which calculates the distances and ranks of each of the candidate CMs of nearness in order when compared to the worst candidate CM valued at 0.0, and the best candidate CM (control) valued at 1.0. Fuzzy set theory and fuzzy membership function as applied to CMs is quite successful in defining the "goodness" or "badness" of a candidate CM. To further elucidate this abstraction, we will discuss an additional example unrelated to CMs.

Figure 3.2 expands upon a simplified example of the membership function with fuzzy sets for "heights". In this circumstance the degree of sharp-edged membership for "tall" is valued at 1.0, while the it is valued at 0.0 for "not tall". However, what about those measures for heights that fall somewhere between "tall" and "not tall"? Such degrees of membership for being somewhere in-between is then a continuous membership (fuzzy membership) function for "tall", and defines the member ship functions for "definitely a tall person" valued at 0.95 and a "really not very tall at all" person valued at 0.30.

3.2 Concept Map Feature Selection

The features extracted from the CM graphs have to be representative of the pre-defined rubric which establishes the ground truth. Two types of features were extracted from the CMs which are as follows: *Content Measure Features* and *Graph Structural Features* [11].

3.2.1 Concept Map Content Measure Features

In this section, we discuss the extracted content measures of cybersecurity CMs, such as the number of nodes, edges, words, characters, hierarchies, crosslinks, and the highest hierarchy.

Number of Concepts

The number of concepts (nodes) is defined as the number of concepts, containing main topic or subtopic concepts. A greater number of concepts found within a CM is presumed to be more

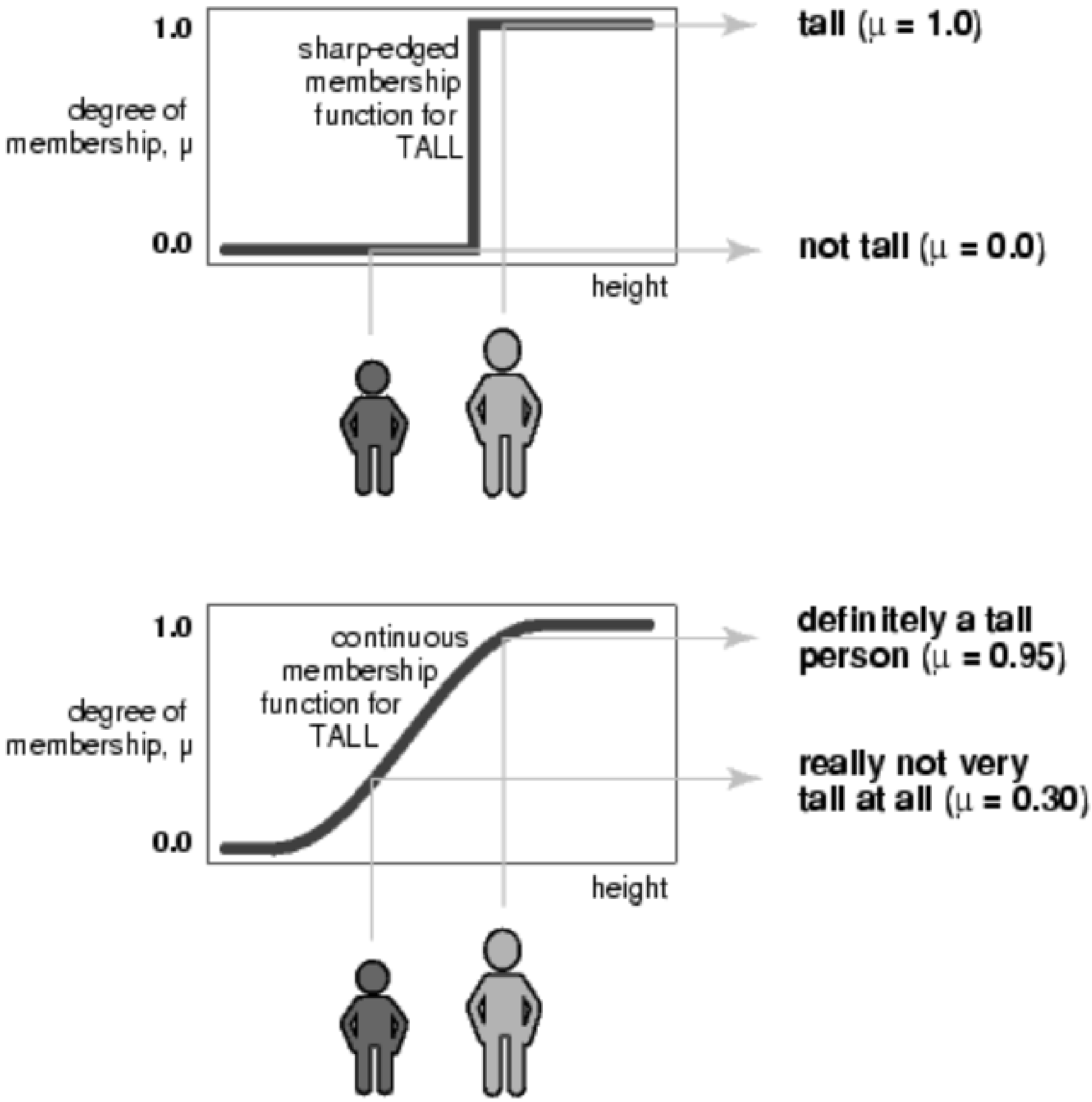

Figure 3.2: Example of Membership Function for "tall" and "not tall". [2].

expansive when compared to a CM with a lesser number of concepts.

Number of Hierarchies

The number of hierarchies is defined as the total number of hierarchy levels. A greater number of hierarchies found with a CM is presumed to be more extensive when compared to a CM with a lesser number of hierarchies.

Highest Hierarchy

The highest hierarchy is defined as the uppermost level of node placement within the CM.

Number of Crosslinks

The number of crosslinks is defined as the total amount of crosslinks between nodes. A greater number of crosslinks found in a CM is presumed to be more interconnected when compared to a CM with a lesser number of crosslinks.

Number of Edges (Linkages)

The number of edges is defined as the number of linkages between a pair of concepts. A greater number of edges (linkages) found within a CM is presumed to be more connected when compared to a CM with a lesser number of edges.

Number of Words per Concept (Average)

The average number of words is defined as the total number of words used on an average in all concepts as well as the linkage phrases. A greater number the average within a CM is presumed to contain a more detailed elaboration of the topic and its subtopics when compared to a CM with a lesser word count.

Number of Characters per Concept (Average)

The number of characters is defined as the average number of characters used in the concepts as well as the linkage phrases. Similarly to the number of words, a greater the average number of characters within a CM is presumed to contain a more detailed explanation of the topic and its subtopics when compared to a CM with a lesser average character count.

3.2.2 Concept Map Graph Structural Features

In this section, we discuss the extracted structural features of CMs.

Distribution of the In-Out Degrees per Concept

The number of in-degrees and out-degrees of a concept is an important feature to represent structural branching of the CM graph. However, the distribution of the number of in-out degrees of a graph is a 1-D set. To minimize the number dimensions of said set, we fitted an exponential distribution to the 1-D set and extracted the mean and the standard deviation of the said set as our features. As the mean and the standard deviation completely characterizes a Normal distribution, we used these two features per graph as our two feature vectors to represent the structural branching of the CM graph. The features follow the following trend: both mean and standard distribution of the distribution are directly proportional to the structural branching of the CM graph.

Degree Centrality (Average)

Degree centrality is defined as the number of links incident upon a node. The degree can be interpreted in terms of the immediate risk of a node for catching whatever is flowing through the network. In the case of a directed network, we usually define two separate measures of degree centrality, namely in-degree and out-degree. Accordingly, in-degree is a count of the number of ties directed to the node and out-degree is the number of ties that the node directs to others.

Closeness Centrality (Average)

Closeness centrality of a node is defined as a measure of centrality in a network, calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes.

Betweenness Centrality (Average)

Betweenness is a centrality measure of a vertex within a graph. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

Cardinality of the Maximal Clique

The cardinality of a set is a measure of the number of elements of the set. Therefore, the cardinality of a maximal clique is defined as a measure of the number of elements of the clique, or subgraph, where the clique of the graph is of maximum possible size.

These features were further used to define the heuristic scoring for the CMs. Each feature represents one dimension of the feature space and a particular CM is represented by a thirteen dimensional point in the feature space.

3.3 Fuzzy Reformulation of the Problem

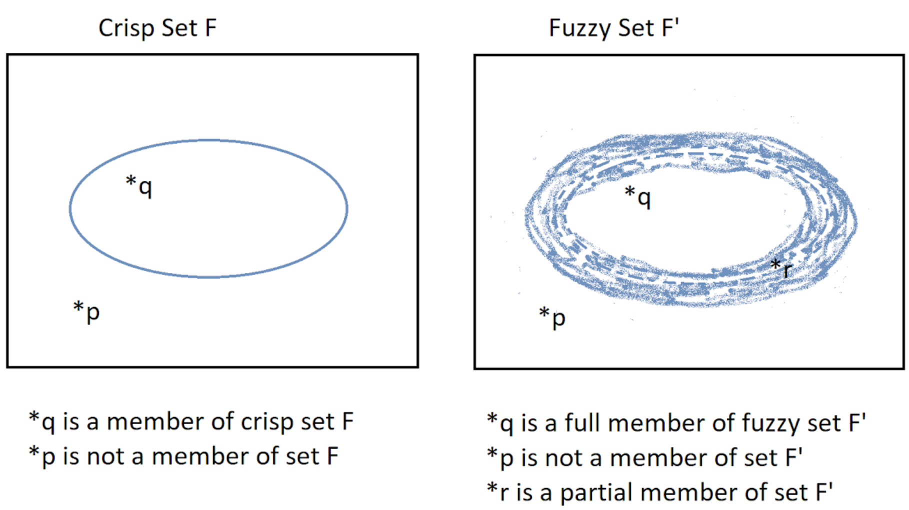

Figure 3.3: Crisp and Fuzzy Sets.

Probability theory is capable of representing only one of several distinct types of uncertainty. As shown in the example in Figure 3.3, when **F'** is a fuzzy set and **r** is a relevant object, the proposition **r is a member of F'** is not necessarily either true or false. But instead, this is true only to some degree, the degree to which **r** is actually a member of **F'**. Thus, the crisp set **F** is defined in such a way as to dichotomize individuals in some given universe of discourse into two groups of members and nonmembers. However, many classification concepts do not exhibit this characteristic. Therefore, a fuzzy set can be defined mathematically by assigning to each possible individual in the universe of discourse a value representing its grade of membership in the fuzzy set [21]. From fuzzy set theory and fuzzy logic, we were able to derive the fuzzy reformulation of the problem.

Since our objective is to score the concept maps (CMs) based on a continuum from a "bad" CM to the model control CM, our problem can be formulated as follows:

Let a CM graph be $CG(N, E)$, where N is the number of nodes, and E is the number of edges. Let the worst CM possible be defined as $CG_0(N, E)$, where $N = 0$ and $E = 0$, and the model control CM be defined as $CG_1(N, E)$ where $N = N_{ControlCM}$ and $E = E_{ControlCM}$. Then, our goal is to find a functional map from the set CG to the interval $[0, y]$, i.e $F(CG(N, E))$ such that $F(CG_0) = x$ and $F(CG_1) = y$, where x and y are the class encoding of CG_0 and CG_1 respectively.

To come up with such a functional map, the first step is to reformulate the problem in terms of fuzzy space and come up with membership functions.

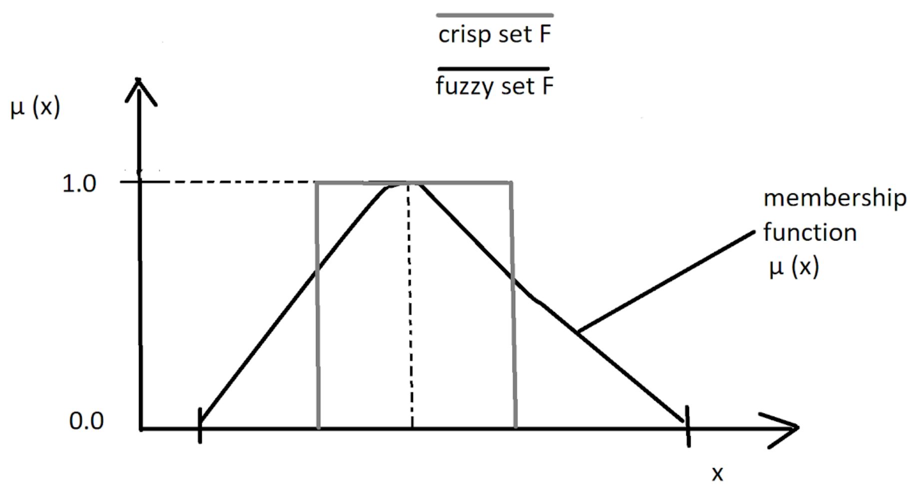

Figure 3.4: Representation of the Fuzzy Membership Function.

3.4 Fuzzy Membership Function

A membership function [12, 22] is a characteristic function in which the values assigned to the elements of the universal set fall within a specified range and indicate the membership grade of their elements in the set. Larger values denote higher degrees of set membership. A set defined by membership functions is a fuzzy set. The most commonly used range of values of membership functions is the unit interval [0, 1]. Reformulating the problem in terms of fuzzy sets, our problem has two sets A and B where $CG_0 \epsilon A$ and $CG_1 \epsilon B$. The membership function of an unknown CG to set A can be defined as follows:

$$\mu_A(CG) = 1 - Norm\|\mathbf{CG} - \mathbf{CG_0}\| \tag{3.1}$$

,where the distance measure is the *Euclidean distance*. Similarly, the membership function of the same CG to set B would be in:

$$\mu_B(CG) = 1 - Norm\|\mathbf{CG} - \mathbf{CG_1}\| \tag{3.2}$$

From the above Equations (3.1) and (3.2), we can see the following: if CG_0 is CG, then $\mu_A(CG)$ is 1 and $\mu_B(CG)$ is 0, and if CG_1 is CG, then $\mu_A(CG)$ is 0 and $\mu_B(CG)$ is 1. This is evident because CG_0 is the same as CG_0 and it definitely belongs to A and the vice-versa.

3.5 Fuzzy Similarity Scaling (FSS) Score

The manner in which we have defined the membership function allows us to now define a functional map $F(CG(N, E))$. Here, let the numerical class encoding of $CG_0(N, E)$ be x and the encoding of $CG_1(N, E)$ be y, where x≥1 and y≥1 (because of the multiplicative nature of the `FSS` scoring function), then for a concept map CG, the functional map $F(CG)$ is:

$$\mathit{F(CG(N,E))} = x * \mu_A(CG) + y * \mu_B(CG) \tag{3.3}$$

We define range of the function for any given CG to be the heuristic `FSS` score for a CM. For any given feature representation of a CM CG, Equation (3.3) gives us with a heuristic continuum measure of where the given CG lies between CG_0 and CG_1. For example, if the `FSS` score of a CG is 1.3, then the CG is qualitatively closer to CG_0, and if it is 1.7, then the CG is qualitatively similar to the model CM CG_1.

Previous attempts at defining a similarity measures have involved subjectively finding similarity based on intersections of the concept sets. However, we believe as such metrics can be incorrect by merely rewording the concepts. We also believe that our heuristic measure provides a more objective way to score the quality of a CM via replacing argumentative measures with quantifiable metrics.

Chapter 4

Implementation Details

Figure 7.1 represents the workflow of the `MAnanA` framework. The first step is to convert a .cmap file to .cxl file. The .cxl files are fed into the parser and the previously specified features are extracted. In our case, we arbitrarily defined the class encoding of set A and B from the previous section as 1 and 2 respectively. Then, values of the membership function is computed for the CM. The `FSS` is then calculated, and the "goodness" of the CM is obtained according to the predetermined range between 1 and 2.

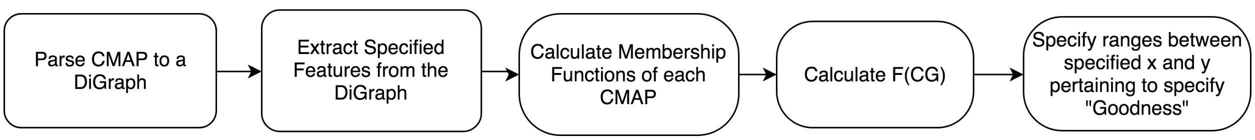

Figure 4.1: Representation of the Workflow of the `MAnanA` Framework.

We implemented the `MAnanA` framework in the Python programming language and utilized the *networkx* and *sklearn* packages. Then the nearest neighbor algorithm is modified for the computation of `FSS` by calculating the distances and ranks of nearness in order, and is described in pseudocode below.

Algorithm 1
K Nearest Neighbor Algorithm. [23]

Classify ($\mathbf{X}$, $\mathbf{Y}$, x)
$\mathbf{X}$ = training data
$\mathbf{Y}$ = class labels of $\mathbf{X}$
x = unknown sample
for $i = 1$ ***to*** m **do**
  Compute distance $d(\mathbf{X}_i, x)$
end
Compute set I containing indices for the k smallest distances $d(\mathbf{X}_i, x)$
return majority label for $\mathbf{Y}_i$ where $i \in I$

Chapter 5

Experiment Setup

This section will discuss the details of the experiment and includes *Data Collection* and *Experiment Procedure.*

5.1 Data Collection

For the participating students enrolled in the course titled "Introduction to Computer Security", *CmapTools* [24] program and instructions for using the diagramming tool for creating CMs were provided. *CmapTools* was used over other diagramming tools for the following reasons: its development by the Florida Institute for Human and Machine Cognition (IHMC); its open availability for download for usage in educational settings; its user-friendly graphical interface for students; and the ability to export the .cmap files as .cxl for automation of analysis. The instructor presented each topic in a format of slide sets and class lectures, where once the slide set topic was presented, the instructor then assigned the students the task of developing a CM for that topic to be completed and submitted within a week. The students were asked to create a separate CM corresponding to the presented slide sets of the lectures for each of the following topics: *Introduction to Computer Security, User Authentication, Cryptographic Tools, Malicious Software*, and *Denial of Service Attacks.*

5.2 Experiment Procedure

The completed CMs are electronically submitted in both .cmap and .pdf files. The .cmap file is the native file format for *CmapTools.* Afterwards, the students' CMs were exported as a .cxl file, a basic XML file that can be parsed to extract features of the CMs. The .cxl files are comprised of such information as the concepts or nodes, the relationships or edges, the labels of the nodes and edges, and the graphical layout of the CM content. The analysis of such types of information is expected to provide an understanding of the knowledge represented by the cybersecurity CMs.

The .cmap files were also graded based on the rubric in table 7.1 out of 35 points in total. Finally, the `FSS` score and scaled rubric scores were plotted together for each main topic as discussed in the results section.

Chapter 6

Establishing the Ground Truth

In this work we created a primary scoring metric to objectify the significance of a CM pertaining to a topic, we required a secondary metric, that has been well established in literature, to justify our scoring metrics usability and significance [25–27].

6.1 Development of Rubric

For this purpose, we developed a rubric to score the overall completeness or "goodness" of a CM, the presence or lack of required concepts, and the existence and correctness of relationships between concepts [11] within the cybersecurity topic domain. This rubric was proposed to assess student concept maps and provided as the ground truth base comparison for the `FSS` scoring. The rubric is comprised of measurements of characteristics that were or were not found in the candidate concept maps. The control CM is expected to contain all relevant and correct conceptual topics, subtopics, and linkages leading to the highest rubric score of 35, such that, the control map meets all qualifications defined in the rubric as seen in Table 7.1. The textual description of each score is now presented in the following section.

Table 6.1: Concept Map Rubric for Establishing Ground Truth

Criteria $\Downarrow$ / Score $\Rightarrow$	Score 5	Score 4	Score 3	Score 2	Score 1
Incorrect Main Topic Concepts	None	≤ 2	≤ 4	≤ 6	≥ 7
Missing Main Topic Concepts	None	≤ 2	≤ 4	≤ 6	≥ 7
Incorrect Linked Main Topics	None	≤ 2	≤ 4	≤ 6	≥ 7
Incorrect Subtopic Concepts	None	≤ 2	≤ 4	≤ 6	≥ 7
Missing Subtopic Concepts	None	≤ 2	≤ 4	≤ 6	≥ 7
Incorrect Linked Subtopics	None	≤ 2	≤ 4	≤ 6	≥ 7
Missing Elaboration of Subtopics	None	≤ 2	≤ 4	≤ 6	≥ 7

6.2 Textual Descriptions of Rubric Scoring

6.2.1 Elaboration of a Score of 5

Mastery in comprehension of the ideas and concepts of the subject matter presented. Displays a complete and well-defined knowledge of interpersonal relationships between the main topic and its subtopics in the most concise and precise manner; includes all of the significant concepts and ideas related to the main topic and its subtopics. Utilizes the relevant linking phrases between the main conceptual topic and subtopics in a concise and precise manner without any duplication; complete and appropriate usage of directional arrows without duplication. Utilizes the knowledge of the main conceptual topic and subtopics with significant usage of relevant and specific examples; complete elaboration of significant key terms.

6.2.2 Elaboration of a Score of 4

Approaching mastery in comprehension of the ideas and concepts of the subject matter presented. Displays mostly a complete and well-defined knowledge of the interpersonal relationships between the main topic and its subtopics in a mostly concise and precise manner; includes a majority of the significant concepts and ideas related to the main topic and its subtopics. Utilizes relevant linking phrases between the main conceptual topic and subtopics in a concise and precise manner without any duplication; mostly complete and appropriate usage of directional arrows without any duplication. Utilizes the knowledge of the main conceptual topic and subtopics with some usage of relevant and specific examples; some elaboration of significant key terms. Presents a relevant title related to main conceptual topic and subtopics; appropriate usage of directional arrows; readable and follow-able formatting; correct spelling and grammar.

6.2.3 Elaboration of a Score of 3

Basic comprehension of the ideas and concepts of the subject matter presented. Displays the minimal knowledge of the interpersonal relationships between the main topic and subtopics in a concise and precise manner; includes the minimal significant concepts and ideas related to the main topic and its subtopics. Utilizes the minimal yet relevant linking phrases between the main topic and subtopics; minimal usage of relevant linking phrases without any duplication; minimal yet appropriate usage of directional arrows without any duplication. Utilizes the knowledge of the main conceptual topic and subtopics with the minimal usage of relevant and specific examples; minimal elaboration of significant key terms. Presents a relevant title related to the main conceptual topic and subtopics; minimal usage of directional arrows; readable and followable formatting; correct spelling and grammar.

6.2.4 Elaboration of a Score of 2

Less than basic comprehension of the ideas and concepts of the subject matter presented. Displays less than the minimal knowledge of the interpersonal relationships between the main topic and its subtopics in a not-concise and imprecise manner; minimal or complete lack of inclusion of significant concepts and ideas related to the main topic and subtopics. Utilizes less than minimal knowledge of the main topic and subtopics with usage of irrelevant, inappropriate, or complete lack of linking phrases and/or any duplication; minimal or incorrect usage of directional arrows and/or any duplication. Utilizes little of the knowledge of the main topic and subtopics with usage of minimal or complete lack of relevant and specific examples; minimal or complete lack of elaboration of significant key terms. Presents a complete lack of a relevant title related to the main topic and subtopics; minimal or complete lack of appropriate directional arrows; lack of readable and/or followable formatting; incorrect spelling and/or grammar.

6.2.5 Elaboration of a Score of 1

Complete lack of any comprehension of the ideas and concepts of the subject matter presented. Displays a complete lack of the minimal knowledge of the interpersonal relationships between the main topic and its subtopics in an inconcise and imprecise manner; complete lack of inclusion of significant concepts and ideas related to the main topic and subtopics. Utilizes none of the knowledge of the main topic and subtopics with usage of irrelevant, inappropriate or complete lack of linking phrases and/or any duplication; minimal or incorrect usage of directional arrows and/or any duplication. Utilizes none of the knowledge of the main topic and subtopics with a complete lack of relevant and specific examples; complete lack of elaboration of significant key terms. Presents a complete lack of a relevant title related to the main conceptual topic and subtopics; complete lack of appropriate directional arrows; complete lack of readable and/or follow able formatting; incorrect spelling and/or grammar.

The textual descriptions of the scores for the ground truth concept map rubric were elaborated upon in order to provide clarification of the characteristics expected in the control CM.

Chapter 7

Evaluation Results

In this section, we provide the results of the experiment. There were about 135 CMs made between five topics. An example set of CMs constituting of the "control CM","good CM", and "bad CM" for the forth topic *Malicious Software* is shown in Figures 7.3, 7.4 and 7.7 respectively. Corresponding values for the features are shown in Table 7.1. The extracted features tables for additional four topics of *Introduction to Computer Security*, *User Authentication*, *Cryptographic Tools*, and *Denial of Service Attacks* follow afterwards.

Table 7.1: Extracted Features for the *Malicious Software* Concept Map Examples of "Control", "Good" or High Score, and "Bad" or Low Score.

Features ⇓ / Concept Maps ⇒	Control	High Score	Low Score
Number of Concepts (Nodes)	241	157	9
Number of Hierarchies	5	22	4
Highest Hierarchy	6	4	2
Number of Crosslinks	0	0	0
Number of Edges (Linkages)	241	156	8
Mean of Degree Histogram	10.9546	6.8260	1.5
Std Dev of Degree Histogram	32.6210	18.8556	2.5
Number of Words per Concept (Avg)	60.3569	38.7236	49.45
Number of Characters per Concept (Avg)	386.0581	249.552	318.52
Degree Centrality (Avg)	0.0084	0.0127	0.2222
Closeness Centrality (Avg)	0.0063	0.00887	0.1296
Betweenness Centrality (Avg)	7.47e-05	9.9308e-05	0.0079
Cardinality of Max Clique	2	2	2

To compare `FSS` based scoring mechanism to the rubric based ground truth, we compare the histogram obtained from the rubrics to the trendline obtained via `FSS`, and perform peak matching. These results are presented in Figures 7.9, 7.11, 7.13, 7.2 and 7.15. We see in the aforementioned diagrams that the trendline obtained from `FSS` is representative of the scores obtained via the rubric, that is, both show an increasing or decreasing trend in the same interval. In some cases, they are even exact like in Figure 7.2. Among the observed results, the one with most difference

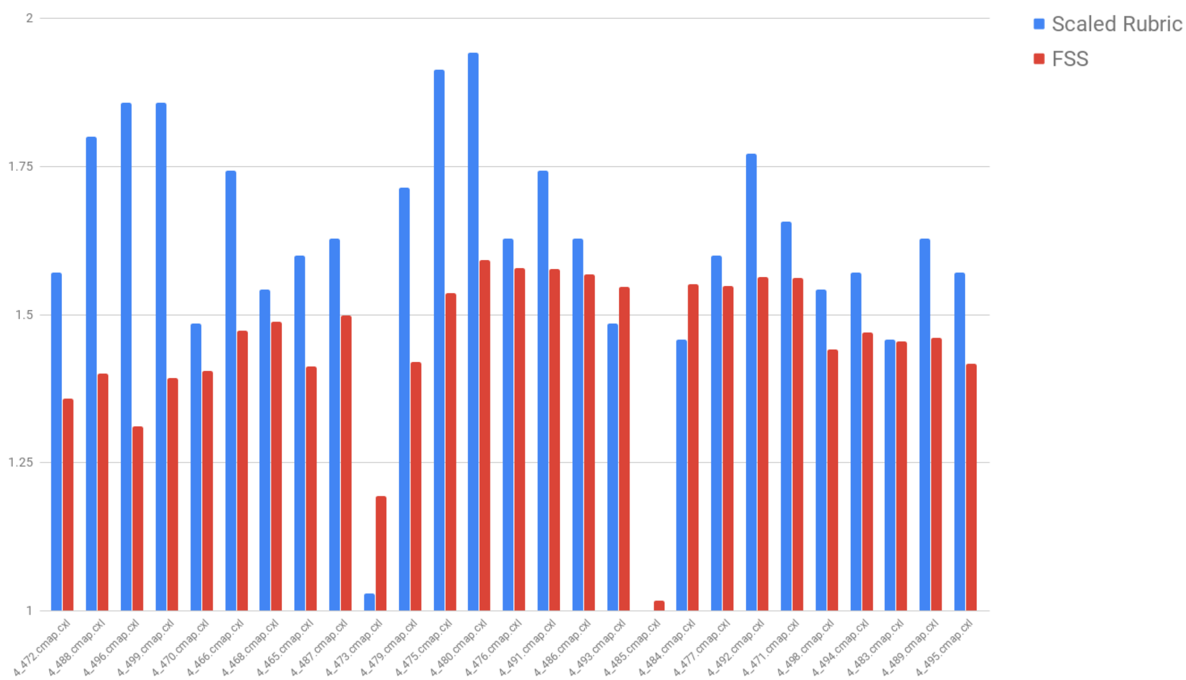

Figure 7.1: *Malicious Software*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) bar graph showing both the plot of the results from each. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind.

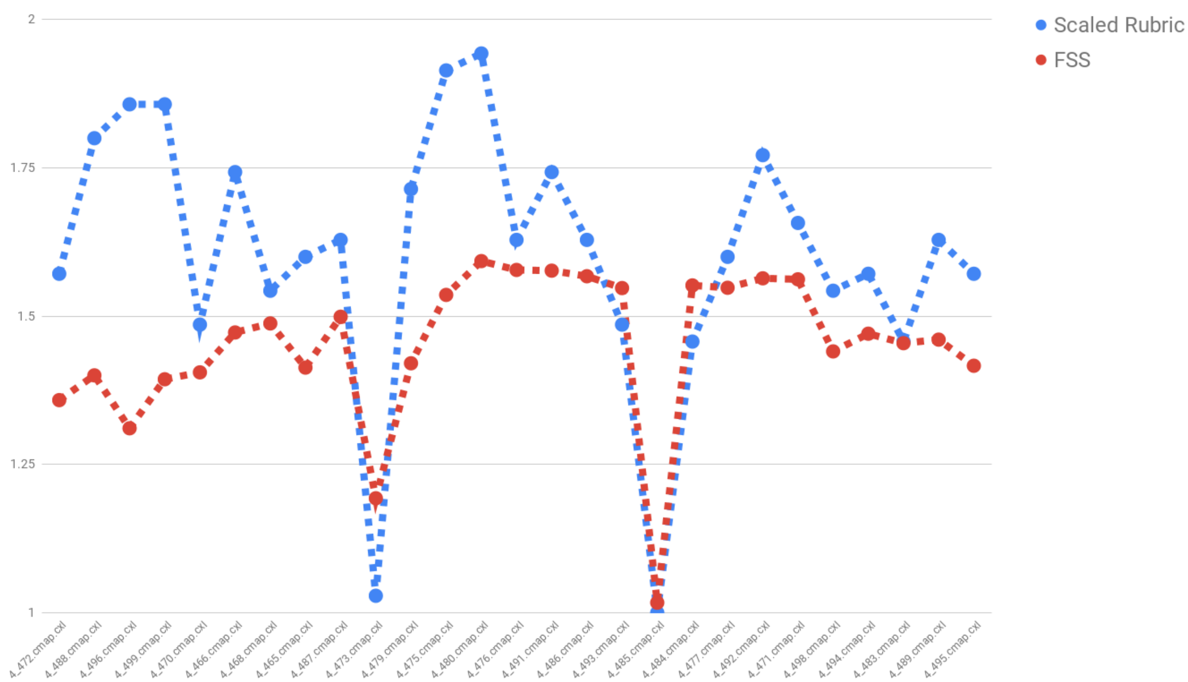

Figure 7.2: *Malicious Software*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) line graph showing peak-matching. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind. Observe the peak-matching for the CMs shown to be empirically "bad".

Figure 7.3: Example of the Control CM for *Malicious Software*. The map is included to display the intricate complexity and well-branched structure, and much elaboration of the designed Control. Zoom into the “Worms” subtopic in Figure 7.5 on the following page.

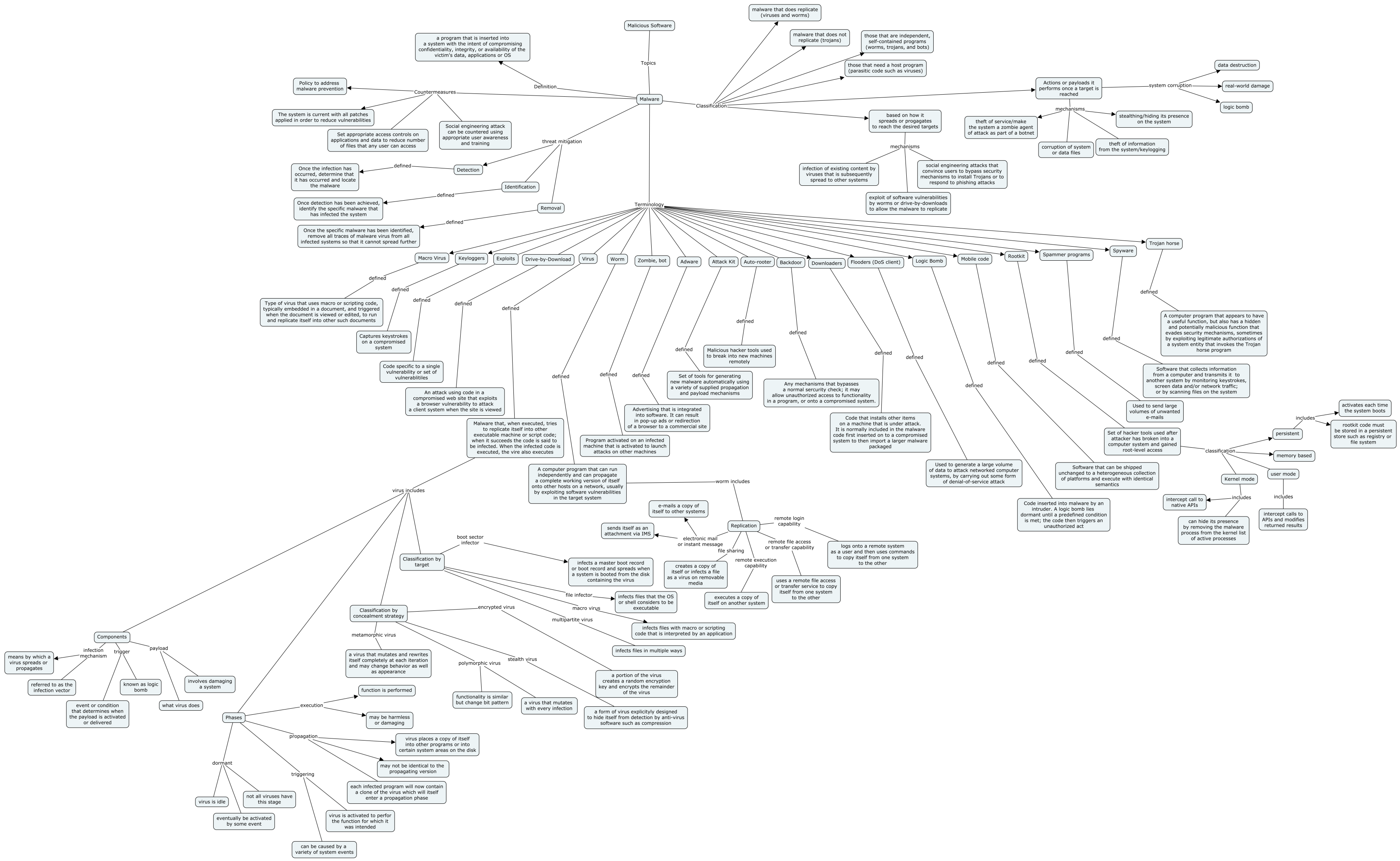

Figure 7.4: Example CM for *Malicious Software* which was rated quite high both by the aforementioned CM Rubric as well as FSS. The map is included to display the complex and well-branched structure of a "good" map. Zoom into the *Worms* subtopic in Figure 7.6 on the following page.

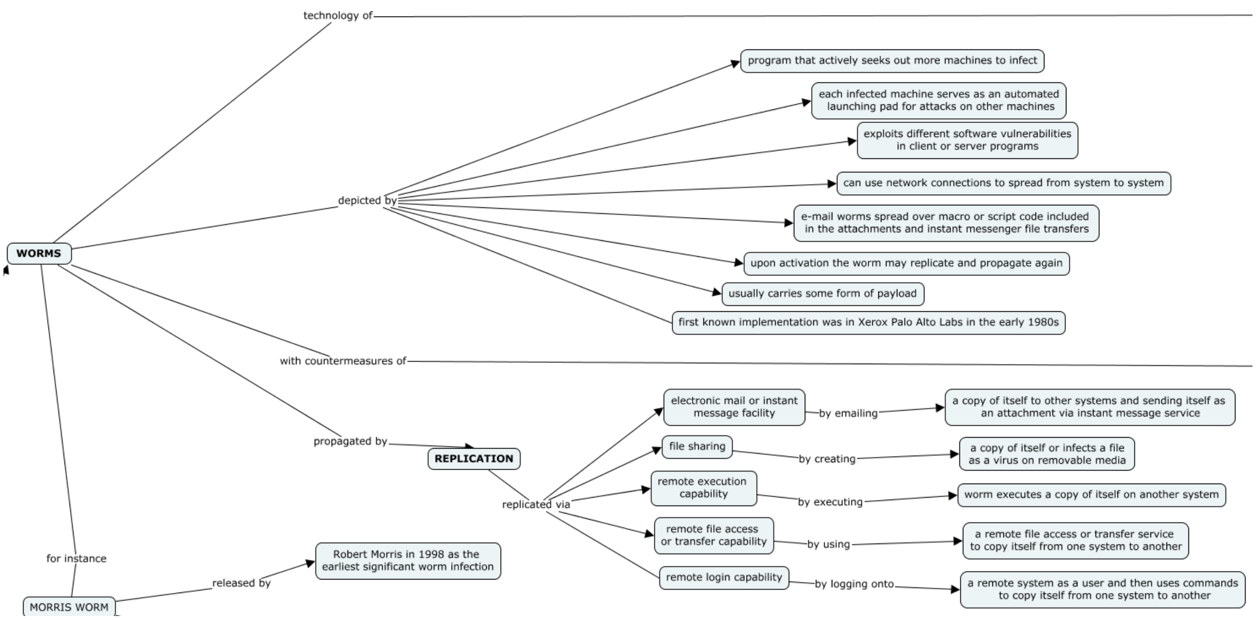

Figure 7.5: Snippet of prior Example CM for *Malicious Software*, subtopic *Worms*, which was rated quite low both by the aforementioned CM Rubric as well as FSS. *Worms* subtopic extends beyond image with further elaboration.

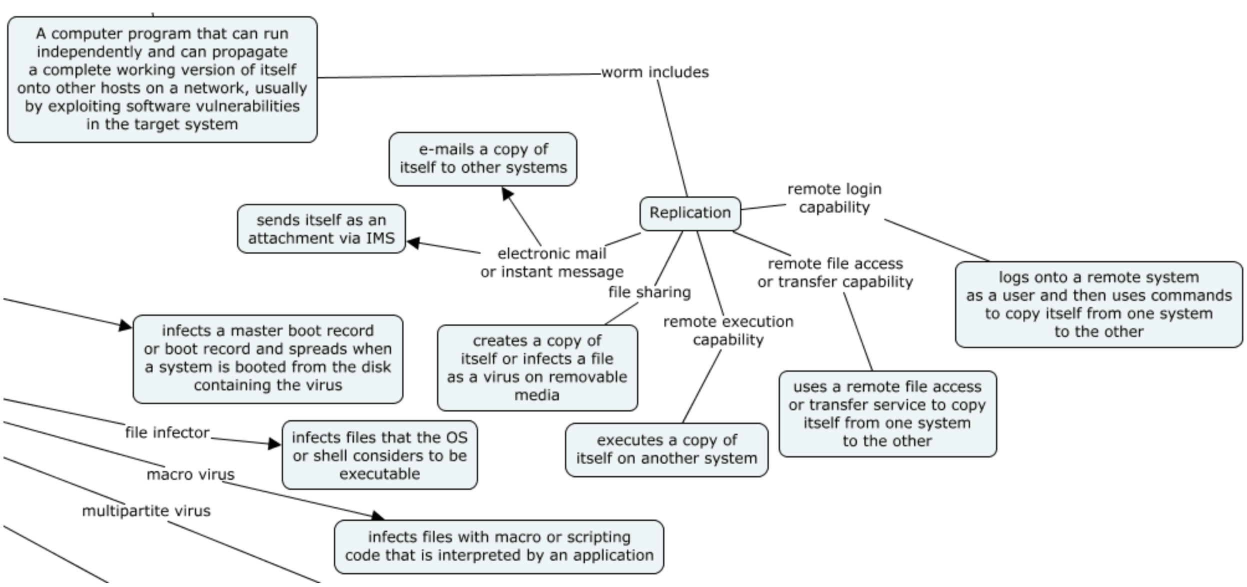

Figure 7.6: Snippet of prior Example CM for *Malicious Software*, subtopic *Worms*, which was rated quite low both by the aforementioned CM Rubric as well as FSS. *Worms* subtopic shown as example of a “good” map.

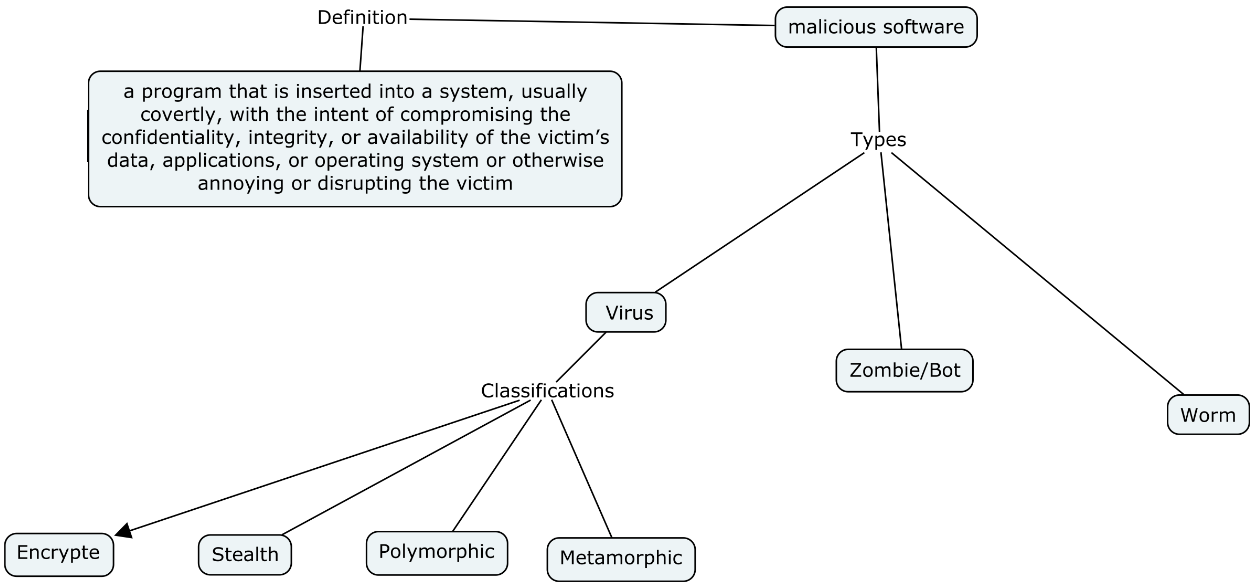

Figure 7.7: Example CM for *Malicious Software*, subtopic *Worms*, which was rated quite low both by the aforementioned CM Rubric as well as FSS. *Worms* subtopic, in lower right corner, shown as example of a "bad" map.

between the ground truth and FSS is the first set in Figure 7.9. We believe that this is because the students did not have a handle at making concept maps at the beginning of the class, and it improved as the class progressed.

Based on the results, we are able to claim that the FSS based scoring methodology we have developed can objectively provide a scoring system to grade and compare CMs.

Table 7.2: Extracted Features for the *Introduction to Computer Security* Concept Map Examples of "Control", "Good" or High Score, and "Bad" or Low Score.

Features ⇓ / Concept Maps ⇒	Control	High Score	Low Score
Number of Concepts (Nodes)	110	55	8
Number of Hierarchies	6	5	1
Highest Hierarchy	6	4	5
Number of Crosslinks	4	0	0
Number of Edges (Linkages)	119	54	12
Mean of Degree Histogram	4.5833	5.5	1.3334
Std Dev of Degree Histogram	13.1146	9.2005	1.2472
Number of Words per Concept (Avg)	51.8455	7.0364	31.1733
Number of Characters per Concept (Avg)	362.4909	49.0364	223.7867
Degree Centrality (Avg)	0.0199	0.0364	0.4286
Closeness Centrality (Avg)	0.0143	0.0253	0.2601
Betweenness Centrality (Avg)	0.000289	0.000743	0.03571
Cardinality of Max Clique	2	2	3

Table 7.3: Extracted Features for the *User Authentication* Concept Map Examples of "Control", "Good" or High Score, and "Bad" or Low Score.

Features ⇓ / Concept Maps ⇒	Control	High Score	Low Score
Number of Concepts (Nodes)	174	76	10
Number of Hierarchies	16	8	1
Highest Hierarchy	7	4	2
Number of Crosslinks	0	0	0
Number of Edges (Linkages)	176	110	9
Mean of Degree Histogram	10.2353	8.4444	1.0
Std Dev of Degree Histogram	25.9309	7.8047	2.6833
Number of Words per Concept (Avg)	37.8563	28.0325	28.7724
Number of Characters per Concept (Avg)	254.9195	192.3008	196.6016
Degree Centrality (Avg)	0.0117	0.0386	0.2
Closeness Centrality (Avg)	0.00837	0.0264	0.1296
Betweenness Centrality (Avg)	0.0001151	0.0005121	0.01111
Cardinality of Max Clique	2	2	2

Table 7.4: Extracted Features for the *Cryptographic Tools* Concept Map Examples of “Control”, “Good” or High Score, and “Bad” or Low Score.

Features ⇓ / Concept Maps ⇒	Control	High Score	Low Score
Number of Concepts (Nodes)	192	110	9
Number of Hierarchies	10	8	2
Highest Hierarchy	7	6	3
Number of Crosslinks	3	0	0
Number of Edges (Linkages)	213	109	8
Mean of Degree Histogram	11.2941	7.8571	1.8
Std Dev of Degree Histogram	27.5228	18.6580	1.7205
Number of Words per Concept (Avg)	53.1198	41.3577	8.9343
Number of Characters per Concept (Avg)	329.2031	257.6496	55.8613
Degree Centrality (Avg)	0.011616	0.018182	0.22222
Closeness Centrality (Avg)	0.008114	0.01289	0.1482
Betweenness Centrality (Avg)	0.0001079	0.000256	0.02778
Cardinality of Max Clique	3	2	2

Table 7.5: Extracted Features for the *Denial of Service Attacks* Concept Map Examples of “Control”, “Good” or High Score, and “Bad” or Low Score.

Features ⇓ / Concept Maps ⇒	Control	High Score	Low Score
Number of Concepts (Nodes)	138	65	23
Number of Hierarchies	11	14	10
Highest Hierarchy	4	3	4
Number of Crosslinks	2	0	0
Number of Edges (Linkages)	144	64	22
Mean of Degree Histogram	11.5	4.3334	2.0909
Std Dev of Degree Histogram	25.3952	12.0314	3.8719
Number of Words per Concept (Avg)	72.5145	46.5983	10.0435
Number of Characters per Concept (Avg)	457.8261	296.1966	63.4783
Degree Centrality (Avg)	0.0152	0.03077	0.0869
Closeness Centrality (Avg)	0.0112	0.0197	0.0551
Betweenness Centrality (Avg)	0.0001373	0.0002518	0.003953
Cardinality of Max Clique	2	2	2

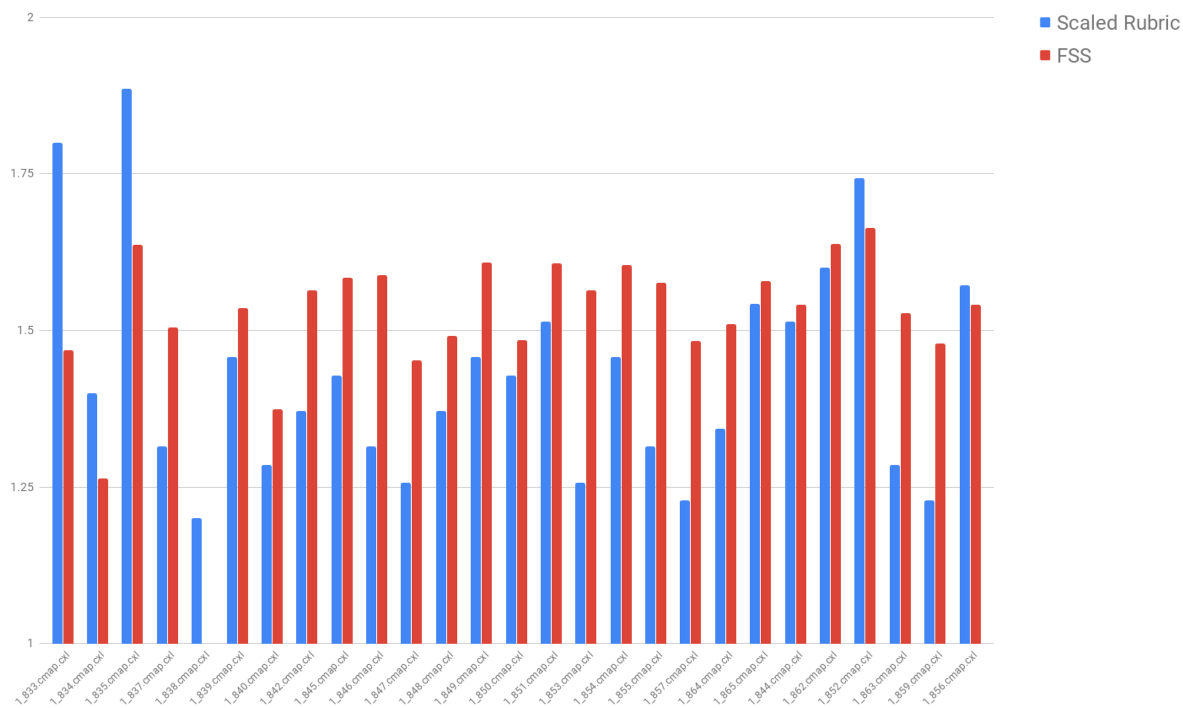

Figure 7.8: *Introduction to Computer Security*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) bar graph showing both the plot of the results from each. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind.

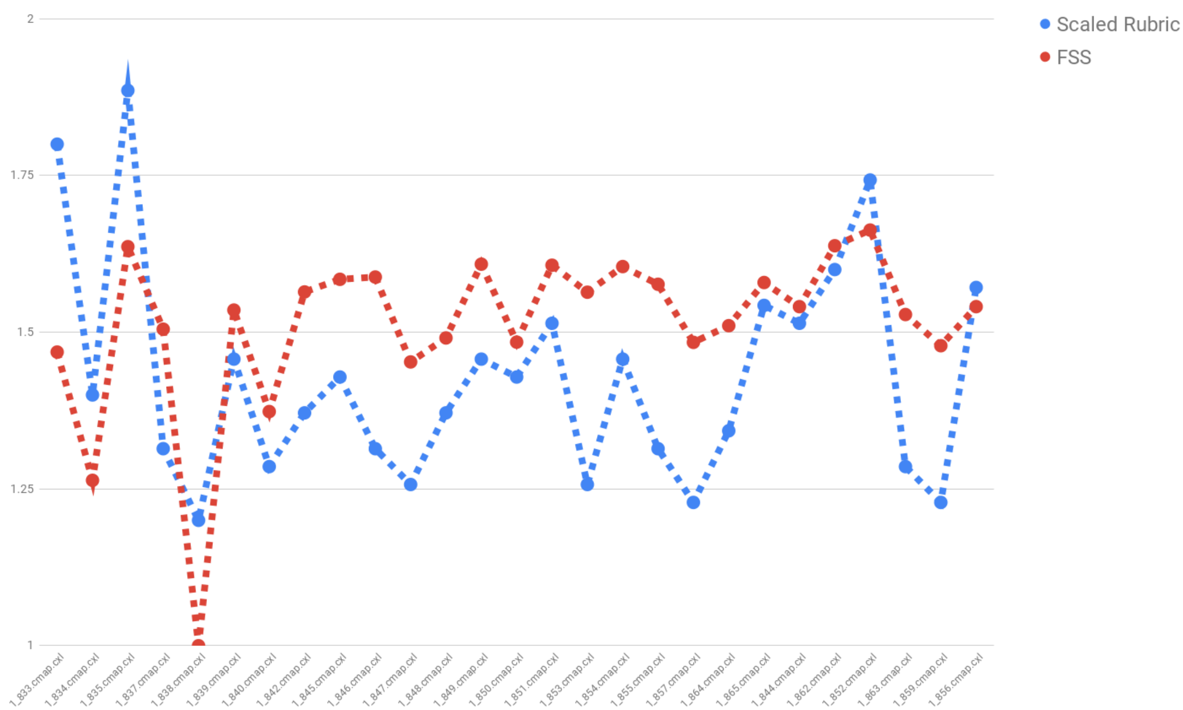

Figure 7.9: *Introduction to Computer Security*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) line graph showing peak-matching. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind. Observe the peak-matching for the CMs shown to be empirically "bad".

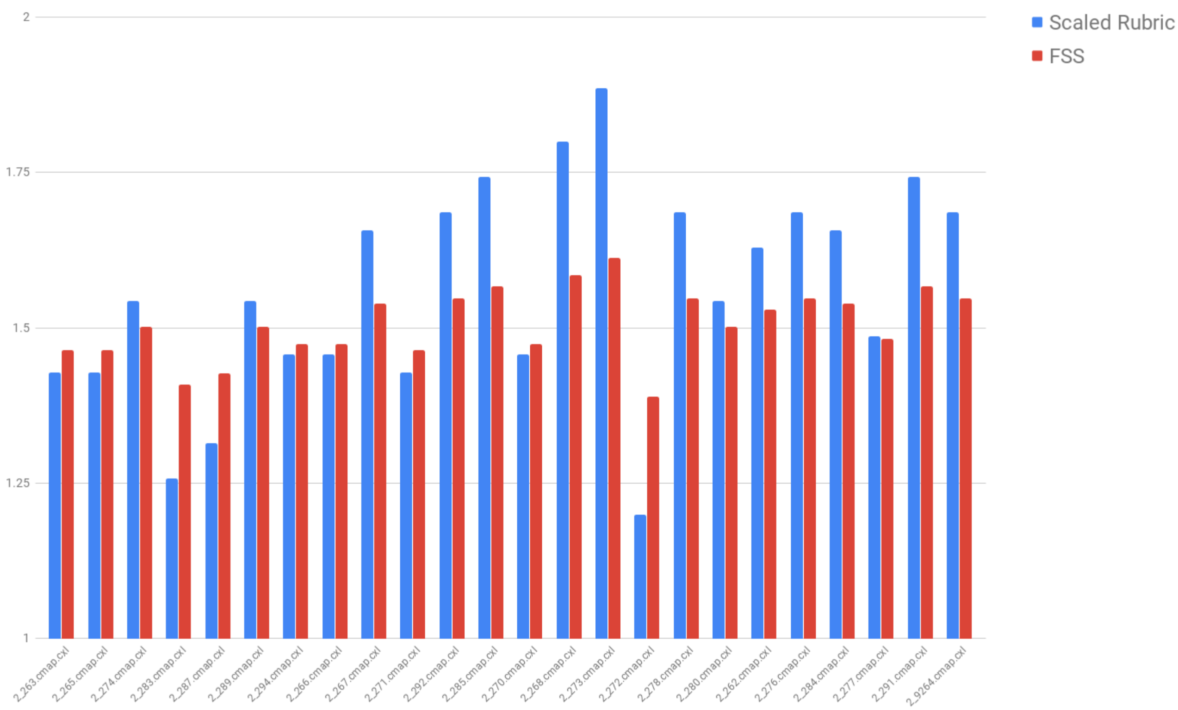

Figure 7.10: *User Authentication*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) bar graph showing both the plot of the results from each. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind.

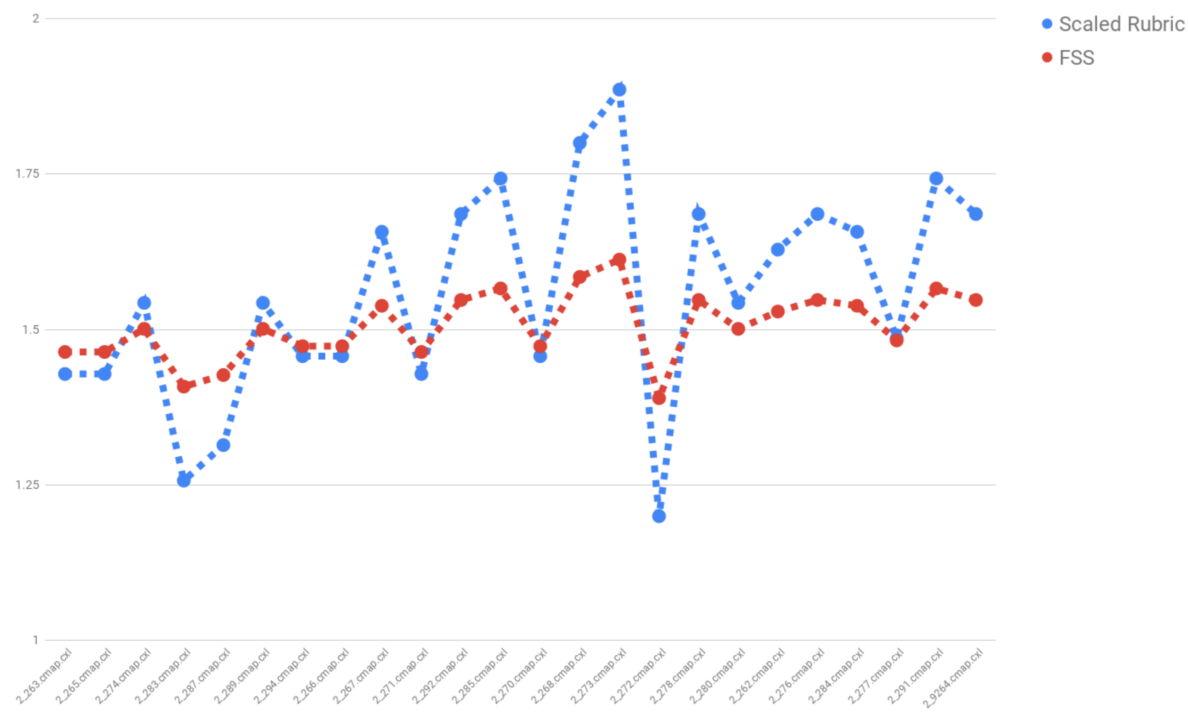

Figure 7.11: *User Authentication*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) line graph showing peak-matching. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind. Observe the peak-matching for the CMs shown to be empirically "bad".

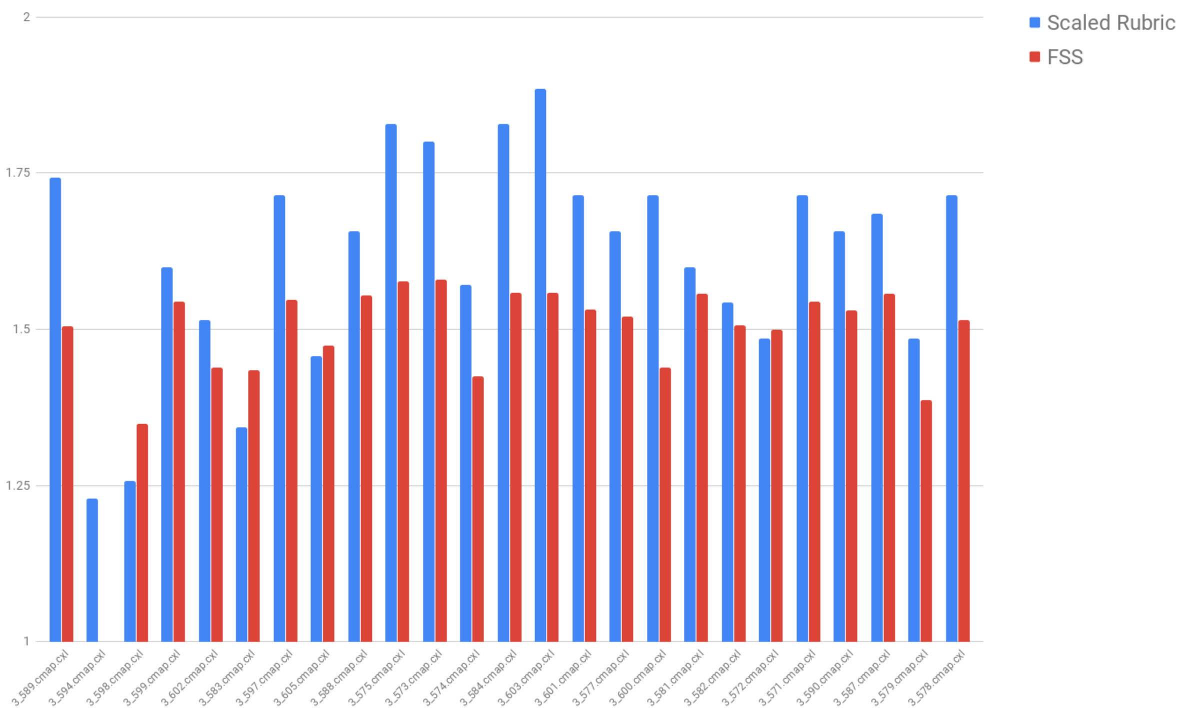

Figure 7.12: *Cryptographic Tools*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) bar graph showing both the plot of the results from each. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind.

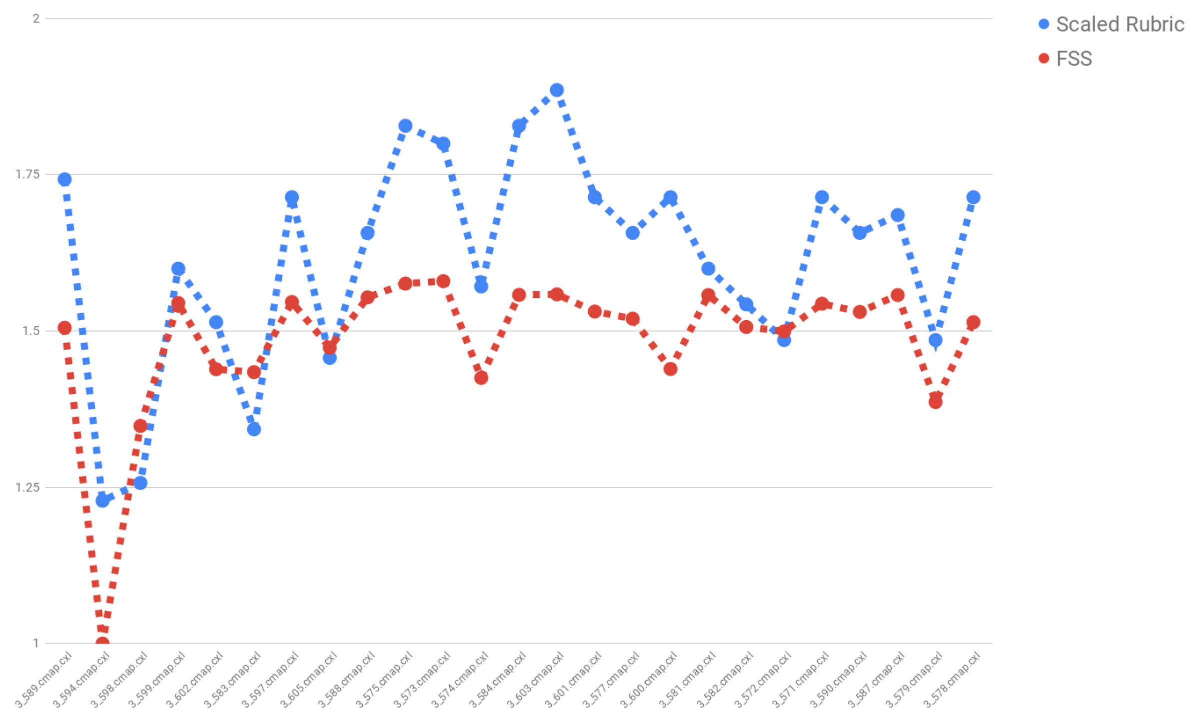

Figure 7.13: *Cryptographic Tools*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) line graph showing peak-matching. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind. Observe the peak-matching for the CMs shown to be empirically "bad".

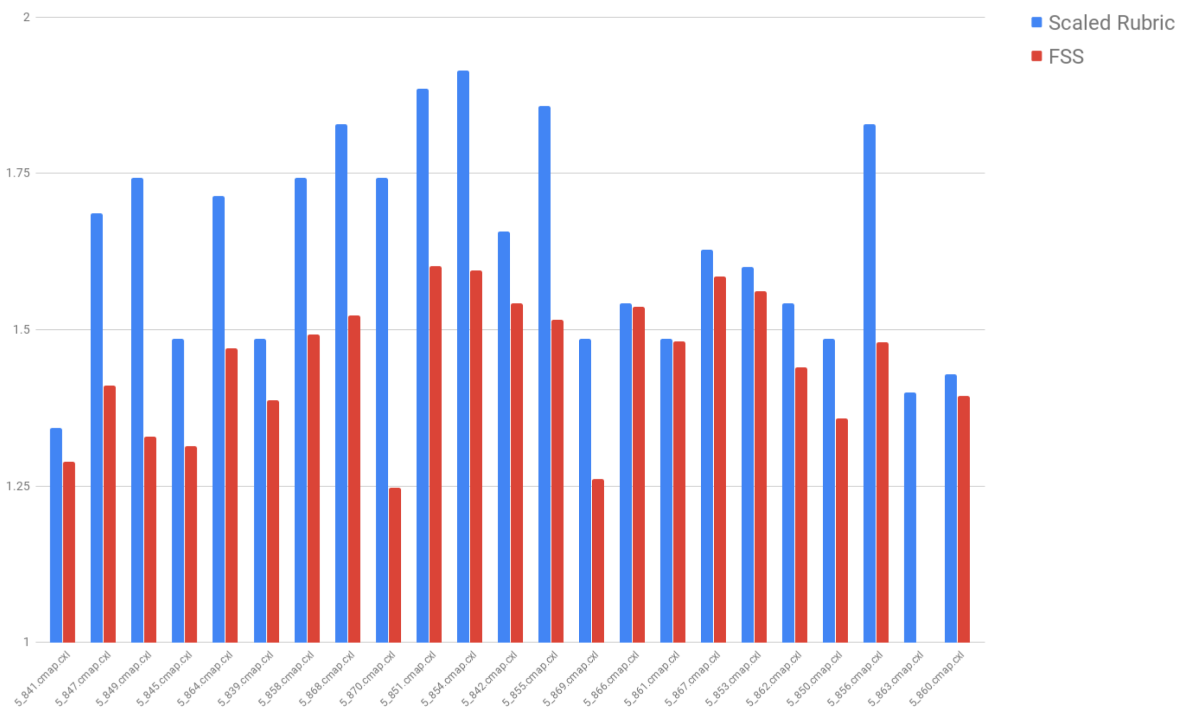

Figure 7.14: *Denial of Service Attacks*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) bar graph showing both the plot of the results from each. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind.

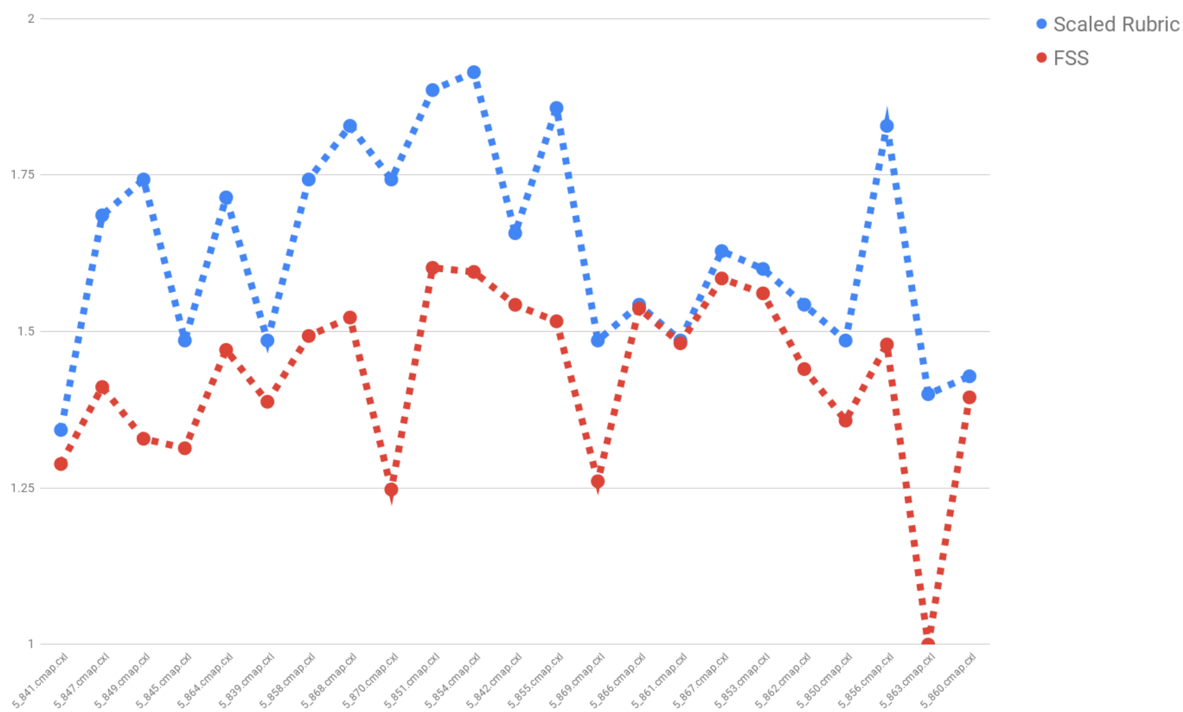

Figure 7.15: *Denial of Service Attacks*: `FSS` scores (in red) vs. ground truth rubric scores (in blue) line graph showing peak-matching. The x-axis lists the candidate CMs, while the y-axis lies within the interval [1,2]. Both the `FSS` scored interval (between 0 and 1), and rubric scores (between 7 and 35) were standardized with the y-axis in mind. Observe the peak-matching for the CMs shown to be empirically "bad".

Chapter 8

Post-Course Survey and Results

8.1 Post-Course Survey

On a side-note, in order to acquire direct CM feedback, students were asked to complete an anonymous survey for the course evaluation at the end of the semester. The CM related questions are as follows:

- On average how much time did you spend on creating a concept map?

  Options: Less than 30 minutes; less than 1 hour; less than 3 hours; less than 5 hours; more than 6 hours.

- Do you read the relevant material from the textbook or slides before developing a concept map?

  Options: Yes or No.

- Developing a concept map helps you understand the course material.

  Circle the number: Strongly Disagree 1 2 3 4 5 Strongly Agree.

- I recommend other instructors use concept maps in their courses.

  Circle the number: Strongly Disagree 1 2 3 4 5 strongly Agree.

- Any brief comments on the concept maps.

This was a way for the students to provide feedback pertaining to the usability of CMs in the classroom.

8.2 Post-Course Survey Results

The results of the anonymous post-survey from the students are itemized below:

- On average how much time did you spend on creating a concept map?

  Student responses as seen in Figure 8.1:

  3.33% Less than 30 minutes

  33.33% Less than 1 hour

  55.3% Less than 3 hours

  6.66% Less than 5 hours

  3.33% More than 6 hours

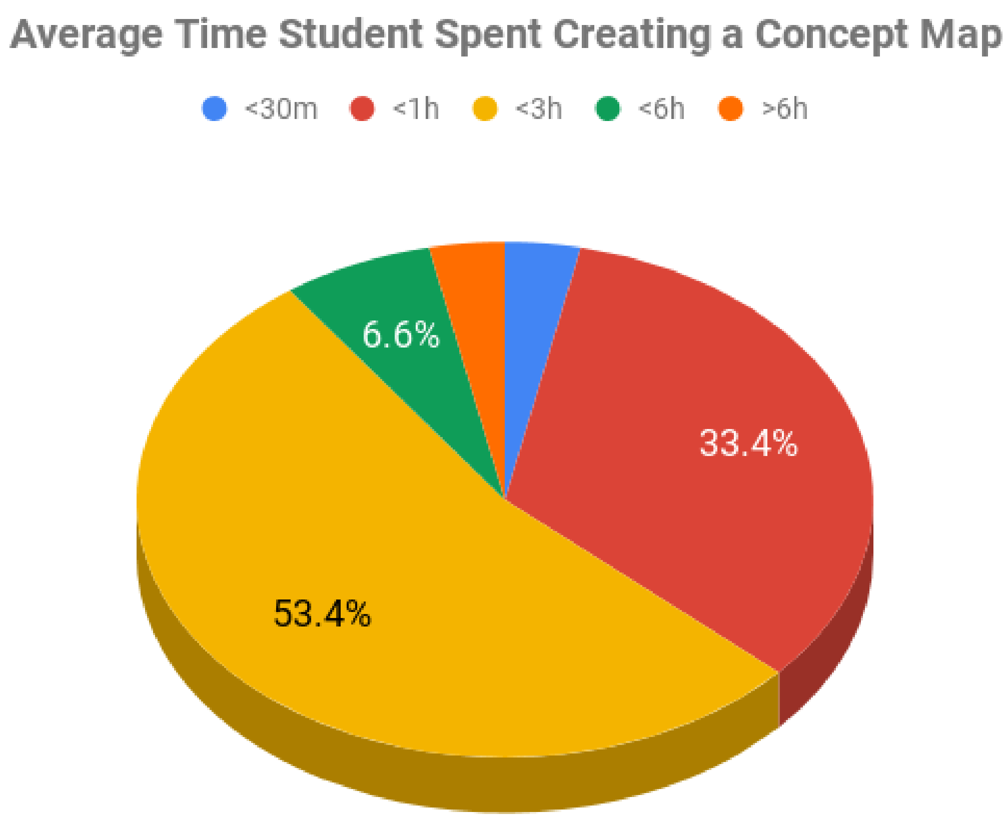

Figure 8.1: Average amount of time spent by student for creation of a concept map plotted by pie chart. As seen above, over half of the students surveyed spent less than 3 hours creating a concept map on average.

- Do you read the relevant material from the textbook or slides before developing a concept map?

  Student responses as seen in Figure 8.2:

  96.67% Yes

  3.33% No

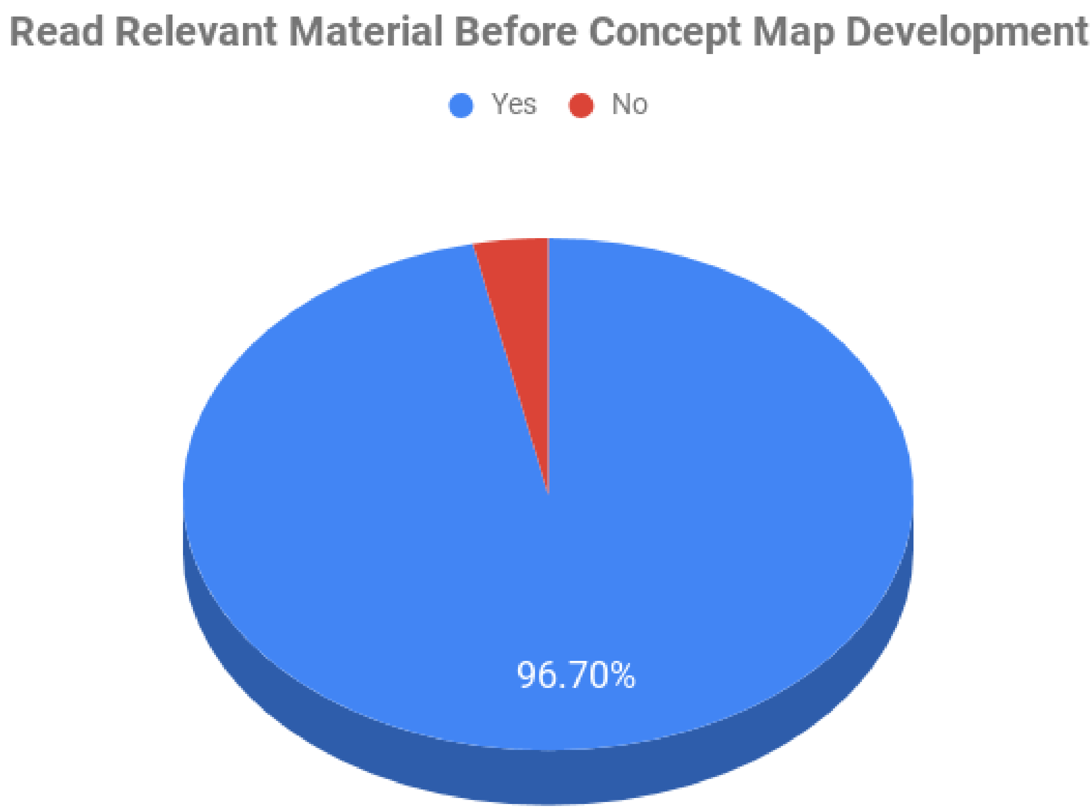

Figure 8.2: The number of students that read the relevant material from the textbook or slides before developing a concept map plotted by pie chart. As seen above, 96.7% of the students surveyed read the relevant material prior.

- Developing a concept map helps you understand the course material.

  Student responses as seen in Figure 8.3:

  23.33% Strongly Agree

  30.0% Agree

  20.0% Neutral

  10.0% Disagree

  16.67% Strongly Disagree

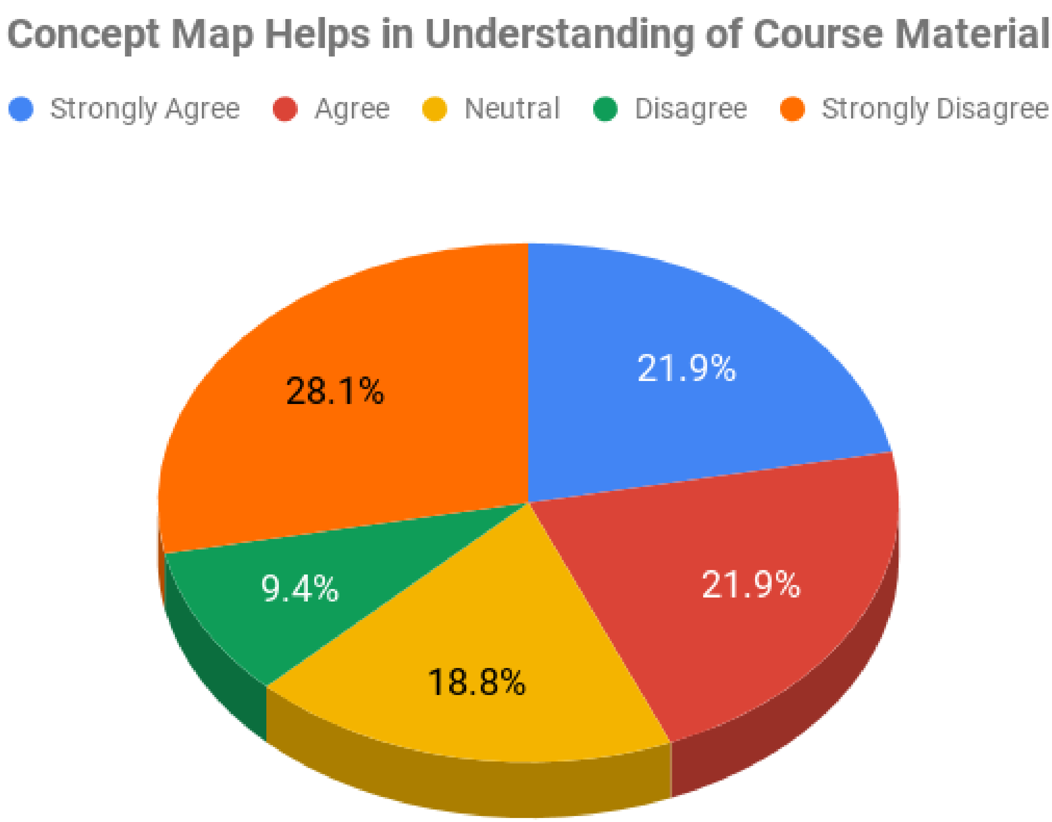

Figure 8.3: Number of students that responded developing a concept map helps in understanding the course material plotted by pie chart. As seen above, 46.3% of the students surveyed agreed that developing a concept map helps in understanding the course material.

- I recommend other instructors use concept maps in their courses.

  Student responses as seen in Figure 8.4:

  13.33% Strongly Agree

  13.33% Agree

  33.33% Neutral

  10.0% Disagree

  30.0% Strongly Disagree

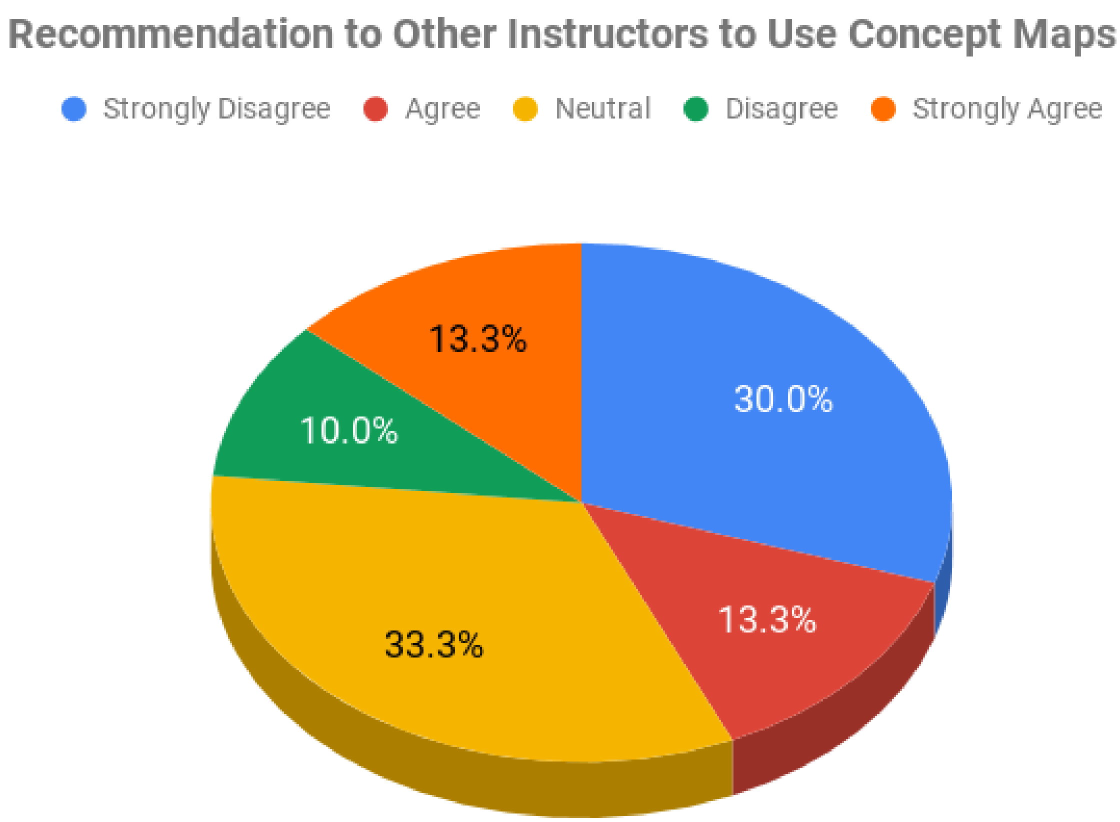

Figure 8.4: Recommendation to other instructors to use concept maps in their courses plotted by pie chart. As seen above, 26.66% of the students surveyed agreed to recommend the use of concept maps to other instructors for their courses.

- Any brief comments on the concept maps.

  Student response as seen in Figure 8.5:

  51.67% no comments

  32.33% positively skewed comments

  16.1% negatively skewed comments

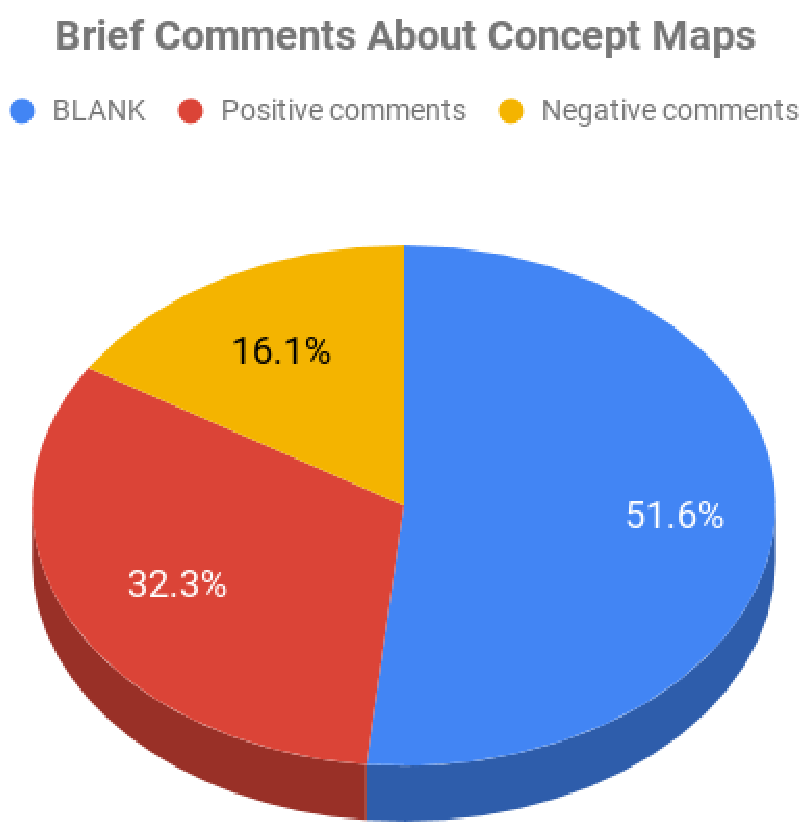

Figure 8.5: Comments by the students on concept maps plotted in pie chart. As seen above, 32.33% of the students had positive comments about concept maps, such as "helpful" and "made me review the material". And 51.67% of the students left no comments.

Chapter 9

Conclusion and Future Work

In this work, we developed an analytic scoring framework, called `MAnanA`, to objectify the scoring mechanism of CMs. Our method is different in the following aspect: it removes the necessity for creating subjective metrics like a rubric which have been traditionally used for the same purpose. To validate the efficacy of our method, we compared the scores via `FSS` and a predefined "ground truth" rubric obtained for around 135 CMs and five control CMs developed by students and educators during the "Introduction to Computer Security" course at UNO. The results made us confident that our scoring framework `MAnanA` provided us with scores that was representative of the quality of the CMs, and could therefore be used to evaluate the "goodness" of the CMs.

One downside of our method is that the CMs that are fed into `MAnanA` have to be preapproved by a subject matter expert who is able to judge relevance. For instance, the framework does not distinguish between a CM for "environment" and a CM for "cybersecurity". However, after a subject matter expert has deemed a CM as relevant, our framework is able to create a score to evaluate a concept maps with respect to a "master CM" with much more objectivity than any previously proposed method. In the future, we hope to explore other distance metrics, and explore similar framework mechanics for quantifying knowledge gain.

Bibliography

[1] Concept map for concept maps (2018).
URL https://www.lucidchart.com/pages/examples/concept-map/bubble-concept-map-temptate

[2] Illustration of membership function (2018).
URL https://edoras.sdsu.edu/doc/matlab/toolbox/fuzzy/fuzzytu3.html

[3] J. D. Novak, D. B. Gowin, Learning how to learn, Cambridge University Press, 1984.

[4] J. Novak, Clarify with concept maps, The science teacher 58 (7) (1991) 44.

[5] J. D. Novak, Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations, Routledge, 2010.

[6] M. Bhatt, I. Ahmed, Z. Lin, Using virtual machine introspection for operating systems security education, in: Proceedings of the 49th ACM Technical Symposium on Computer Science Education, ACM, 2018, pp. 396–401.

[7] W. Johnson, I. Ahmed, V. Roussev, C. B. Lee, Peer instruction for digital forensics, in: 2017 {USENIX} Workshop on Advances in Security Education ({ASE} 17), USENIX, 2017.

[8] I. Ahmed, V. Roussev, Peer instruction teaching methodology for cybersecurity education, IEEE Security & Privacy 16 (4).

[9] W. E. Johnson, A. Luzader, I. Ahmed, V. Roussev, G. G. R. III, C. B. Lee, Development of peer instruction questions for cybersecurity education, in: 2016 USENIX Workshop on Advances in Security Education (ASE 16), USENIX Association, Austin, TX, 2016.
URL https://www.usenix.org/conference/ase16/workshop-program/presentation/johnson

[10] J. D. Novak, A. J. Cañas, The theory underlying concept maps and how to construct and use them.

[11] W. Wei, K.-B. Yue, Using concept maps to assess students' meaningful learning in is curriculum, in: Proceedings of the EDSIG Conference ISSN, Vol. 2473, 2016, p. 3857.

[12] W.-L. Gau, D. J. Buehrer, Vague sets, IEEE transactions on systems, man, and cybernetics 23 (2) (1993) 610–614.

[13] H.-J. Zimmermann, Fuzzy set theory, Wiley Interdisciplinary Reviews: Computational Statistics 2 (3) (2010) 317–332.

[14] M. A. Erceg, Metric spaces in fuzzy set theory, Journal of Mathematical Analysis and Applications 69 (1) (1979) 205–230.

[15] R. Kruse, J. E. Gebhardt, F. Klowon, Foundations of fuzzy systems, John Wiley & Sons, Inc., 1994.

[16] J. R. McClure, P. E. Bell, Effects of an environmental education-related sts approach instruction on cognitive structures of preservice science teachers.

[17] J. R. McClure, B. Sonak, H. K. Suen, Concept map assessment of classroom learning: Reliability, validity, and logistical practicality, Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching 36 (4) (1999) 475–492.

[18] A. Mühling, Aggregating concept map data to investigate the knowledge of beginning cs students, Computer Science Education 26 (2-3) (2016) 176–191.

[19] J. Keppens, D. Hay, Concept map assessment for teaching computer programming, Computer Science Education 18 (1) (2008) 31–42.

[20] D. B. Leake, A. G. Maguitman, A. J. Cañas, Assessing conceptual similarity to support concept mapping., in: FLAIRS Conference, 2002, pp. 168–172.

[21] G. Klir, B. Yuan, Fuzzy sets and fuzzy logic, Vol. 4, Prentice hall New Jersey, 1995.

[22] S. Haykin, Neural networks: a comprehensive foundation, Prentice Hall PTR, 1994.

[23] B. Tay, J. K. Hyun, S. Oh, A machine learning approach for specification of spinal cord injuries using fractional anisotropy values obtained from diffusion tensor images 2014 (2014) 276589.

[24] A. J. Cañas, G. Hill, R. Carff, N. Suri, J. Lott, G. Gómez, T. C. Eskridge, M. Arroyo, R. Carvajal, Cmaptools: A knowledge modeling and sharing environment.

[25] N. L. Miller, A. J. Cañas, A semantic scoring rubric for concept maps: design and reliability.

[26] A. J. Cañas, J. D. Novak, P. Reiska, How good is my concept map? am i a good cmapper?, Knowledge Management & E-Learning: An International Journal (KM&EL) 7 (1) (2015) 6–19.

[27] J. Sánchez, A. Cañas, J. Novak, Concept map: a strategy for enhancing reading comprehension in english as l2, CMC 2010 (2010) 29.

Vita

S. E. Blake Gatto is a long time denizen of New Orleans. She obtained dual degrees of B.A. in Psychology and B.S. in Nursing (Biology minor) from the University of South Alabama. Some time afterwards, she began attending the University of New Orleans with the intent of obtaining a M.S. degree in Computer Science under the Information Assurance concentration.