

# Towards a Better Graphlet-based Mind Map Metric for Automating Student Feedback

Peter Jamieson

Electrical and Computer Engineering

Miami University

jamiespa@miamioh.edu

Jeff Eaton

Electrical and Computer Engineering

Miami University

eatonjw2@miamioh.edu

## Abstract

In this work, we design a new metric that helps analyze the difference between a student's mind map and an expert's mind map. The goal is to determine if student's are learning and to provide them with automatic feedback on what they are not understanding. In previous work, we used various graph analysis metrics applied to the mind map to see if there is evidence of learning, and in this work, we combine ideas from two of the most successful of these metrics by creating a new tool that checks if small sub-graphs exist in both a student and the criterion map (an experts mind map). By analyzing the results of these matches, we create a global metric that we then compare to our previous metrics and find that this new metric has similar behavior. This is positive since this metric provides a means for more interesting feedback to students.

## 1 Introduction

In this paper, we evaluate a new mind map analysis metric that compares an experts mind map (called the criterion map) to a students map to evaluate how similar the two maps are. In previous work, we evaluated graph metrics such as graph density, node degree, Relative Graphlet Frequency (RGF)-distance<sup>1</sup>, and an edge to edge math metric<sup>2</sup> that we created. The later two metrics, when applied in a longitudinal experiment, showed promising results in capturing student improvement.

The context of this research is the idea of a computer-based system that allows students to test their understanding of terminology in a particular subject or field and receive immediate feedback on how they compare to experts.

This work focuses on graphlets, which are small graphs of with 3 to 5 nodes (see Figure 2). The RGF-distance metric, in our previous study, uses a count of each type of graphlet which then can be compared with the criterion maps graphlet count to determine if graphs are similar, and this technique was developed to compare proteins with computational efficiency, where proteins are significantly larger graphs compared to our mind maps. Mind maps, which are a connection of

terms with lines, are small graphs and it is possible to directly compare all graphlets (see if a graphlet in the criterion map is also in the student's map).

To determine our new direct matched graphlet based metric, we analyzed the same data from our the 2 years of longitudinal based experiments previously studied in 2011 and 2012. We built a tool that allowed us to extract and analyze data on which graphlets had sufficient frequency to be useful for student feedback, and we create a new metric based on this analysis. In the results, we find that this metric has similar behavior to both RGF-distance and our edge based match metric where improvement of a student's mind maps is observed over the period of a semester course. This new metric is better than RGF-distance since it compares actual matched graphlets allowing us to use the presence and absence of these graphlets as direct feedback to the student. Additionally, similar to our previous studies, there is no correlation between grades and any of our metrics.

The rest of this paper is organized as follows: section 2 describes some of the related work. Section 3 describes our new metric and provides insight on what this metric used on existing data is showing, and section 4 provides a comparison of our new metric with previous metrics for our 2012 and 2011 experimental data. Finally, section 5 concludes the paper and provides discussion on directions for this research.

## 2 Background

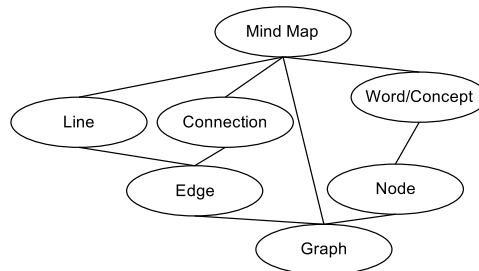


Figure 1: Example of a mind map on the relationship between mind maps and graphs

Mind maps<sup>3</sup> are a visual representation that is used in a number of settings including a Class Assessment Techniques (CATs)<sup>4</sup> which allows teachers to provide feedback on student understanding in class. Figure 1 shows an example mind map that expresses the author's understanding of mind maps and how they relate to mathematical graphs. The words/concepts that are in a mind map are the nodes of a graph (circled bubbles), and the connecting lines between these words are edges of a graph.

Our focus is to analyze mind maps treating them as graphs. Our approach compares student's maps to a criterion map, and Ruiz *et. al.*<sup>5</sup> called this type of scoring *comparison with a criterion map*. There are other scoring mechanisms called *score map*, which uses basic graph analysis techniques such as counting edges and nodes, and *hybrid model*, that combines comparison with criterion and score map. Herl *et. al.*<sup>6</sup> calls mind maps that are restricted in construction as *closed*,

which means the words and concepts are limited, and *open* maps are unrestricted. Our current work focuses on closed construction.

## 2.1 Previous Metrics

Table 1: A summary of our previous findings.

Metric	Shows Improvement
Graph Density Based	No
Average Degree Based	No
Direct Edge Matching	Yes
RGF_distance <sup>1</sup>	Yes

In our previous work<sup>7</sup> and<sup>2</sup> we focused on experimenting if different graph metrics would show if there was an improvement in student’s understanding as a semester progressed. This was tested by seeing if their mind maps were more similar to the criterion map under the assumption that the student was learning throughout the semester. Previously, we examined 4 metrics. In Table 1 we list these metrics and our findings noting that no metric had any correlation to student grades. In column 1, we name the metric, and in column 2, we state whether the metric seems to be showing improvement. For details on these metrics please see<sup>2</sup>.

The first two metrics in Table 1 are global measurements that showed no pattern, and we will not discuss them further. For the other two metrics, we will provide some background on them.

The direct edge metric, which we call *GranularSimilarity* is an edge by edge and node by node comparison between the student mind map and the criterion map. Since the nodes in both graphs are identified uniquely by a label (the word that the node represents is the unique label), we can compare the two graphs and record statistics about the missing edges (*MissE*), missing nodes (*MissN*), extra edges (*ExtraE*), and matching edges (*MatchE*) for the student map compared to the criterion map. If the two graphs are the same, then these statistics will show this since all edges will match and nothing will be missing. We combine these statistics in the following equation:

$$GranularSimilarity = \frac{MatchE}{MissN + ExtraE + MatchE} \quad (1)$$

This will produce a number that has the value between 0 and 1 where the closer the value is to 1 means that there are less missing and incorrect edges.

RGF-distance is the other metric which starts with the idea of graphlets. Graphlets, formally, are “a connected network with a small number of nodes”<sup>1</sup> and these small graphs are non-isomorphic induced subgraphs of a larger graph. Figure 2 shows all the graphlets of size 2, 3, and 4.

The power of the graphlet is how it can be used to analyze a graph. The procedure developed by Przulj *et. al.*<sup>1</sup> is to search for all graphlets of size 3, 4, and 5 in a given graph. Based on the count of each type of graphlet, a signature can be constructed in the form (g0, g1, g2, g3, g4, g5, g6, g7, g8, ..., g28, g29), where g1 is number of the first type of graphlet of size 3 shown in figure 2 and

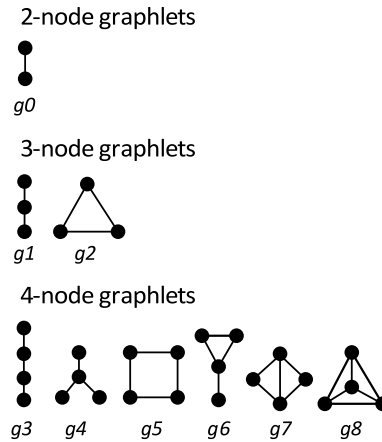


Figure 2: Graphlets of size 2, 3, and 4

$g_{29}$  is the count for the last graphlet of size 5. This signature can be compared to another graphs signature to get a measure of similarity, and Przulj *et. al.* used their technique to compare graphs representing biological structures such as proteins.

RGF-distance is a measure of the difference in frequency of graphlets of  $g_1, g_2, g_3, \dots, g_{28}$ , and  $g_{29}$  appearing in the two graphs being compared. A detailed equation is presented in Przulj *et. al.*<sup>1</sup> and the reader can find the details on the calculation of RGF-distance. As this metric decreases towards 1 the more similar the two graphs are. Note that this measure is an approximation of similarity that can be computed quickly for large graphs. Mind maps are small graphs which allows us to propose a graphlet by graphlet matching.

### 3 An Improved Feedback Metric

Our hypothesis is that based on the set of graphlets used in the RGF-distance metric, certain graphlets relate well to learning, and these graphlets in particular assembled as a new metric will provide better automatic feedback for students. Additionally, because our graphs are small the computational cost of doing direct comparison is feasible.

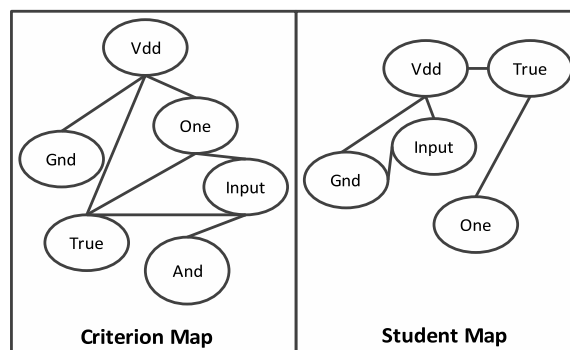


Figure 3: Example Mind Maps (left = criterion, right = student)

To understand what we are proposing for our new metric, take look at figure 3, where the mind map on the left is the criterion map (expert's map) and the one on the right is a student. Our *GranularSimilarity* metric is a one-to-one comparison and basically finds whether one term is related to another term. From this type of comparison, a student could be provided with comments such as: *You are missing the connection between "Vdd" and "One"*. This is an important relationship for a student of digital design to grasp, but deeper relationships can not be captured by this one-to-one comparison. For example, the triangle connecting "one", "true", and "Vdd" is an important relationship that a graphlet based metric would capture in comparison.

We hypothesize that there are some graphlets that relate well to learning. This includes triangles (g2) and fully connected graphs of size 4 (g8) and 5 (g29) which express deeper relationships that are key to learning. Note that some of these ideas were cited in Kinchin and Hay<sup>8</sup>, as described earlier, where they used some of these shapes as matched by a human evaluator. From this basic premise, we will create a new metric that does direct matching of graphlets, but first we need to see what type of matches we have in our existing data to speculate on a metric. Note, however, that the data may provide insights that do not match with our hypothesis - as is the case.

### 3.1 Analysis based on Matched Graphlets

We have built a tool that matches graphlets in the criterion map to graphlets in the students map. This tool not only allows us to compute our graphlet based match metric, but provides us with data to tailor such a metric.

The data set that we ran this tool on was the mind maps created in courses taught in 2011 and 2012 teaching digital system design (Miami University's ECE 287). In those classes, a set of 20 terms was used as an in class activity where students had ten minutes to create a mind map from the provided terms. This activity was done longitudinally by running this activity 3 times over the semester (at the start of the course - pre, after exam 1 - examI, and after exam 2 - examII).

The activity under experimentation is a small part of this course. The students are assessed based on their successful completion of 8 labs, and their performance on 3 exams, 10 quizzes, and 1 major design project. Students meet for two 80 minute lectures during the semester and one 110 minute lab per week. A typical lecture includes presentation, practicing problems, and discussing the material.

Figure 4 shows a histogram bar plot of all of the graphlets that match between a particular student over pre, examI, and examII. This was done for all the students so that we could analyze the data and determine which graphlets are matching (between student and criterion map) and might make sense for assessing learning and providing feedback. Note that g0 graphlets are the same as our *GranularSimilarity* metric.

Some things that can be observed from Figure 4, noting that it represents the data for only 1 student, includes:

- g0, g1, g3, and g9 are all lines of 2, 3, 4, and 5 nodes. These, from a hypothetical learning perspective, might be of minor interest unless something like causality or temporal characteristics are included (which they are not in our case), however, matching of a student

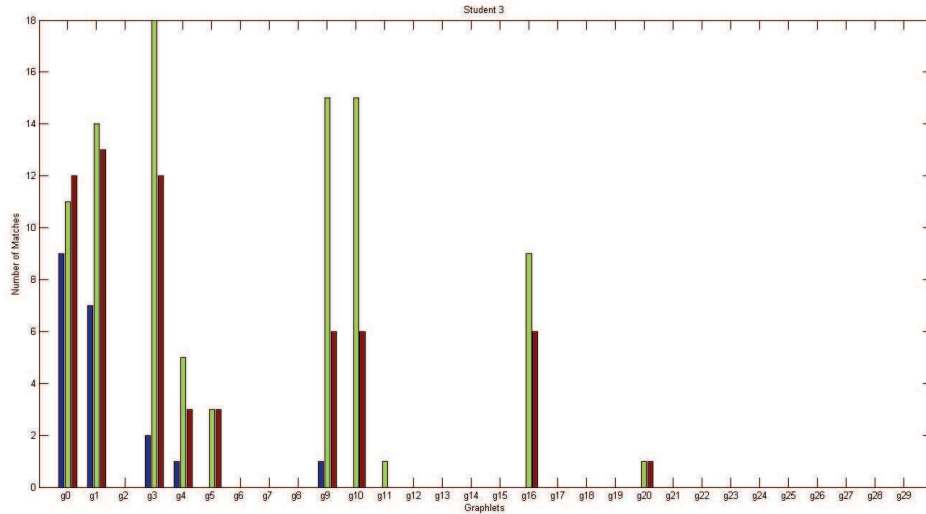


Figure 4: Histogram for matched graphlets of a random student over the longitudinal study in 2011.

to the criterion map suggests that these patterns exist and the student is making the same connections. Note that g0 isn't necessarily the largest of these numbers, which might be assumed. This is because of permutations in a fully connected graph can make many more graphlets of g1, g3 etc. than g0 graphlets.

- The graphlets matched as g16 are squares with a single edges emerging from one point one that square. Again, there's very little from an educational hypothesis that says these links are of value, but if the student is matching the criterion map there is significant impact.

In general, our early hypothesis was that fully connected graphs (g2, g8, g29) were of significant importance to matching between the criterion and the student. However, from this single data point (figure 4, these graphlets are not being matched. Instead, it appears that any complex matching, even if they don't seem like important learning based relationships, are of importance. We might make the following statement, matched complexity in a mind map as observed through graphlets with 5 nodes or less provides feedback to the learner of improvement in their understanding beyond one-to-one matching.

Looking at all the data, we make the following more global statements:

1. There are graphlets that matched between student and criterion maps in such large quantities that they provide little insight on the performance of the student by dominating the existence of other graphlets. These include g3, g9, g10, and g16. However, the absence of these might be of interest.
2. There are graphlets that are matched in very small quantities and we chose to leave them out of the metric. These include g4, g6, g11, g12, g13, g14, g17, and g19. The decision to not use these was based on if the total number of these matched graphlets is less than the number of students in the class then don't include it.
3. The graphlets g21 through g29 are rarely matched, but we hypothesize the existence of them is significant due to their complexity.

### 3.2 Graphlet based match-metric

The general form of our graphlet based match-metric is:

$$GraphletMatch = \frac{\sum_{i \in M} \frac{n_{sg_i}}{n_{G_{g_i}}}}{g_{total}} \quad (2)$$

where  $g_i$  is a graphlet type from  $g_0$  to  $g_{29}$ ,  $M$  is the set of included graphlets in the metric (as in a selection of the graphlets that have meaning derived from the previous section),  $n_{sg_i}$  is the total number of matched metrics in of that type of graphlet,  $n_{G_{g_i}}$  is the total number of that type of graphlet in the criterion map, and  $g_{total}$  is the total number of graphlets (as in  $\sum_{g_i} n_{G_{g_i}}$ ).

This metric will produce a number between 0 and 1 where as the number approaches 1, the student's mind map is closer to the criterion map.

The key question, is what is the set  $M$  that best measures improvement. The reality, however, is such a set is impossible to determine without some other metric of goodness, and instead, we choose to create this set using the data from 2011 and evaluate it relative to both 2011 and 2012. We argue that this produces a better metric than RGF-distance since RGF-distance just produces a signature from a graph that is compared to the signature of the criterion map (this is done because RGF-distance is a computationally cheaper comparison for graphs with 1000s of nodes). Instead, this new metric is an attempt to capture matching graphlets including the labels in the nodes. This has two advantages. First, if the metric improves over the longitudinal study, it is a more accurate estimation of improvement since it is based on graphlets that exist in both the student's mind map and the criterion map. Second, it is possible to identify missing graphlets with this tool as feedback (noting, however, these missing graphlets are numerous and how to report these to the student is left as future work).

From the analysis in the previous section, we created the following set  $M$  based on the 2011 data where

$$M = 0, 2, 5, 7, 8, 15, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29. \quad (3)$$

This includes  $g_0$ , which is one to one mapping, and therefore, has some elements of the our *GranularSimilarity* metric. We will show how this metric performs compared to RGF-distance and the *GranularSimilarity* metrics in the following section.

## 4 Results

The key question is if the new proposed *GraphletMatch* metric performs similar to RGF-distance and the *GranularSimilarity* metric. If this is true, then we can argue that this new metric eliminates the need for the other two as it defines real graphlet matches (RGF-distance does not) and it captures some of the *GranularSimilarity* metric ( $g_0$  graphlets). The only missing aspect is the idea of missing nodes and extra edges that the *GranularSimilarity* metric does capture.

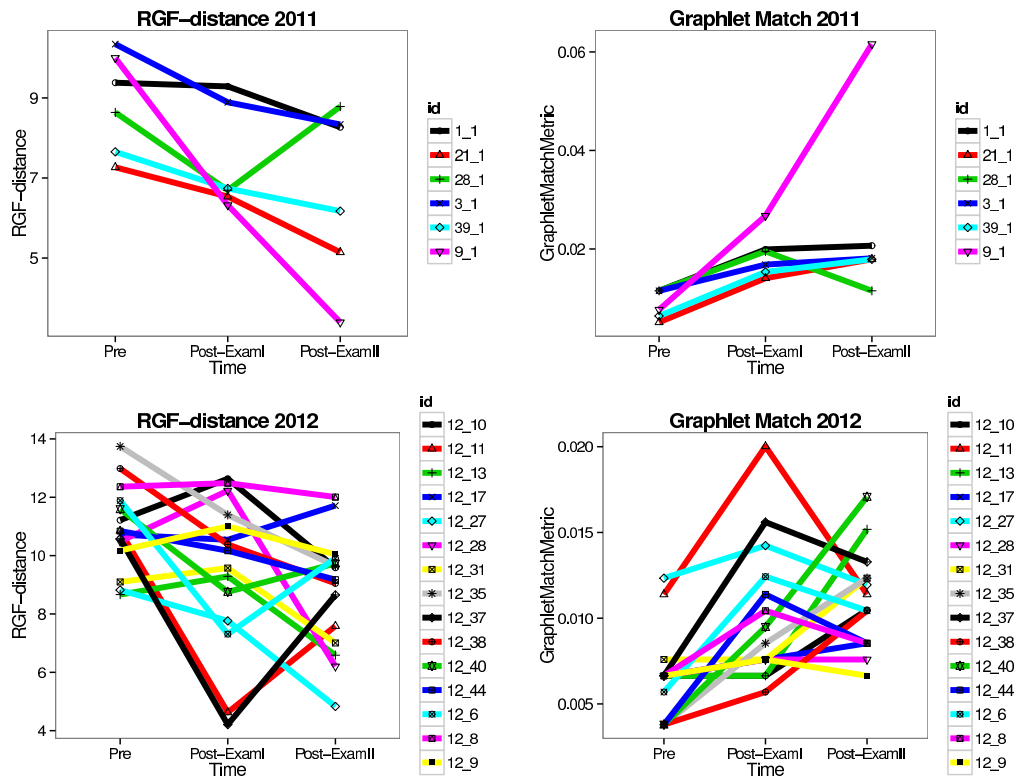


Figure 5: Results for students in the 80 to 90 range for RGF metric and our new metric.

Figure 5 shows a comparison of the RGF-distance metric and the match metric in both 2011 and 2012 data set. This is not the complete set of students, and the subset is students in the grade range 80 to 90 percent who did all three mind map activities. Even though this is a subset of the data, these trends are evident in all the students, but the data is too much to easily view. We remind the reader that the assumption made for this work is, "students are learning over the semester", and therefore, a metric that shows learning will also show improvement over the longitudinal study. For RGF-distance, the metric will move downwards towards 1 as the student map and criterion map become more similar, and for the *GraphletMatch* metric the value will move upwards towards 1 where 0 reflects no matching.

From this figure, it appears that our new metric has a similar behavior to RGF-distance. As noted in our previous work<sup>2</sup>, in many cases student's seem to be performing better after exam I then exam II. We have no reason why this is the case, but we are performing additional experiments to see if we can determine why this is happening. Broadly, it appears that the *GraphletMatch* metric is as good as RGF-distance with the added benefit of being a true matching of graphlets as opposed to RGF-distance's measure of approximate structure.

Figure 6 shows a similar comparison as previous but with the *GranularSimilarity* metric and the new match metric in both 2011 and 2012 data set. This data is only for students in the grade range 80 to 90 percent who did all three mind map activities. In this case, both metrics show improvement as they tend towards 1 with 0 representing no match.



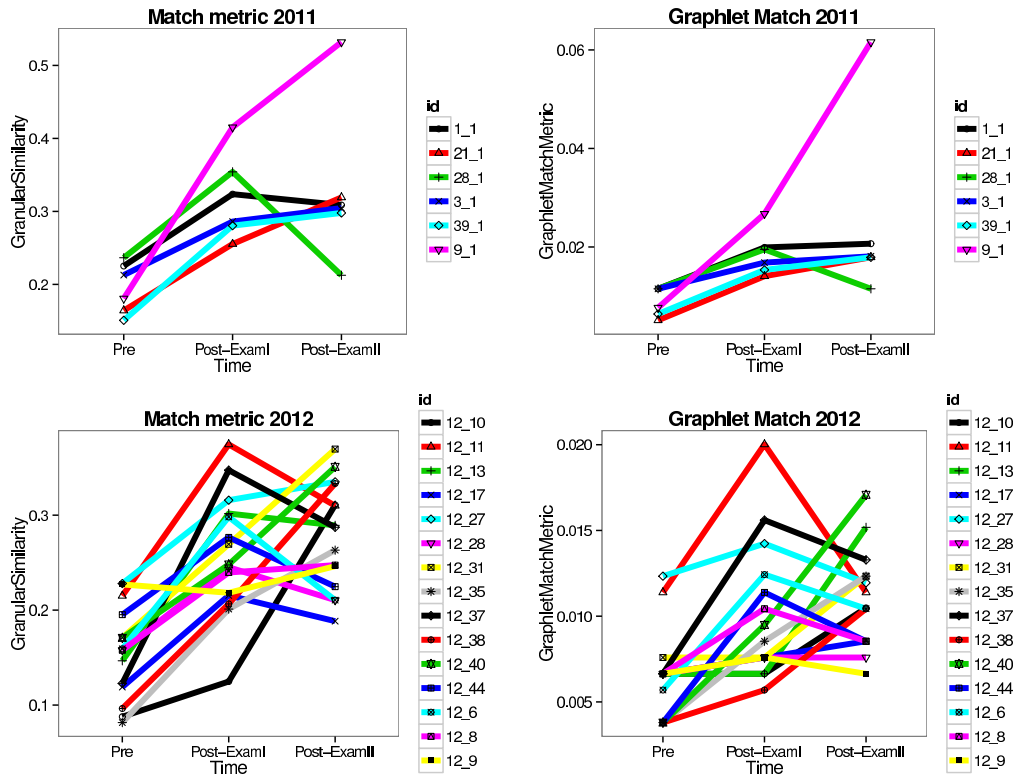


Figure 6: Results for students in the 80 to 90 range for the *GranularSimilarity* metric and our new metric.

In this comparison, we see very similar results as above where both metrics show similar trending of evidence of student improvement. There are a few cases where the new metric behaves slightly differently than the *GranularSimilarity*. For example, some of the students metrics change in terms of the function having a different discrete convex or concave lines (where convex is defined as the slope of the line from point 1 to point 2 being less than the slope of the lone from point 2 to point 3). For example, student 9\_1 shows this in the metrics across 2011 as the magenta line changes. Also, in some cases, the end metric (in post-ExamII) changes from better than the pre measurement to worse than pre in cases such as student 12\_9 in the 2012 metrics. These differences are small, but lead to questions such as how relevant is the numerical value of the metric as opposed to general trending of the curves. At this point, we do not attempt to postulate on these questions, because we have a bigger question that we believe must now be answered (as we will discuss in the conclusion and discussion).

Overall, our key question of whether this new metric provides similar results to our two previous best metrics has been answered with yes, and this new metric is good based on similar behavior. The deeper questions of what does this mean are still open, and in relation to this, researchers might be asking how do the metrics relate to course performance as represented by grades. In all of our experiments, including this one, we have checked for statistical significance as a correlation between grades and our metrics, and as of present, we have found no correlation. Instead, the exercise provides a means to evaluate students on their understanding of terms and

provide automatic means of feedback on where they are performing poorly.

## 5 Conclusion and Discussion

In this paper, we explored the idea of a metric that measures student learning based on analyzing their mind maps compared to an expert mind map. The overall goal of this work is to find a way to provide students with automatic feedback on their understanding of terms within a particular field. In previous work, we created and used measures of graph structure (graphlets, degree, and density) and a metric that looked directly at what was a match and what was missing from the student's mind map on a one-to-one ratio. In this work, we added a new metric that takes ideas from graphlets and our one-to-one matching to determine if measurements of this type were feasible, called *GraphletMatch*.

After analyzing the types of graphlets that are matched between student and the criterion maps, we created a new metric. We evaluated our students in the years 2011 and 2012 with this new metric and compared it to RGF-distance and *GranularSimilarity* metrics to see if we observed similar trends. This was the case and this new metric will allow us to give better feedback to the students based on more complex connected graphs.

In parallel with this work, we have been doing other experiments to improve this research. In particular, we have been exploring if the results from other fields of learning, such as speech pathology and political science, have similar results. This allows us to increase our data set and to understand if these methods are more generally applicable to a range of courses and under what conditions. Our second direction, which we have learned might be the most important part of this work, is how should the criterion map be created? Even though our results seem to show positive results, we suspect that the criterion map is fundamental in this process. For example, if the terms in a field are highly connected (larger degree) then does this make it easier or harder for a student. Also, how does the timing of when the terms are presented in the course impact the longitudinal study. Our plan is to use this new metric in these studies and to try and answer these questions for the community.

The data used in this study are available at:

[www.users.miamioh.edu/jamiespa/DATA\\_SETS/](http://www.users.miamioh.edu/jamiespa/DATA_SETS/) with the "ASEE15" in the name.

## References

- [1] Natasa Przulj, Derek G. Corneil, and Igor Jurisica. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.
- [2] P. Jamieson. More graph comparison techniques on mind maps to provide students with feedback. In *Frontiers in Education Conference (FIE), 2013*, oct. 2013. URL [http://www.users.muohio.edu/jamiespa/html\\_papers/fie\\_13.pdf](http://www.users.muohio.edu/jamiespa/html_papers/fie_13.pdf).
- [3] T.H. Anderson. *Study skills and learning strategies*. Center for the Study of Reading, University of Illinois at Urbana-Champaign, 1978. URL <http://books.google.com/books?id=JNHUjwEACAAJ>.

- [4] Thomas Angelo and K. Cross. *Classroom Assessment Techniques - A Handbook for College Teachers*. Jossey-Bass, 2003.
- [5] M. Ruiz-Primo, R. Shavelson, and S. Schultz. On the validity of concept-map-based assessment interpretations: An experimental testing the assumption of hierarchical concept maps in science. Technical report, University of California, 1997. URL <http://research.cse.ucla.edu/Reports/TECH455.PDF>.
- [6] H. E. Herl, H. F. O'Neil, G. K. W. K. Chung, and J. Schacter. Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior*, 15(3-4): 315–333, May 1999. URL [http://dx.doi.org/10.1016/S0747-5632\(99\)00026-6](http://dx.doi.org/10.1016/S0747-5632(99)00026-6).
- [7] P. Jamieson. Using modern graph analysis techniques on mind maps to help quantify learning. In *Frontiers in Education Conference (FIE), 2012*, oct. 2012. URL [http://www.users.muohio.edu/jamiespa/html\\_papers/fie\\_12.pdf](http://www.users.muohio.edu/jamiespa/html_papers/fie_12.pdf).
- [8] I. M. Kinchin, D. B. Hay, and A. Adams. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Journal of Educational Research*, 42:43–57, 2001. URL <http://www.personal.psu.edu/kmol178/blogs/kmorourke/qualitative%20approach%20to%20concept%20map%20analysis.pdf>.