

More Graph Comparison Techniques on Mind Maps to Provide Students with Feedback

Peter Jamieson

Department of Electrical and

Computer Engineering

Miami University

Oxford, Ohio 45056

Email: jamiespa@miamioh.edu

Abstract—One of the limiting aspects in education research is the techniques available for determining if a student has learned something. In this work, the goal is to extend our exploration of how mind maps can be automatically analyzed using their graph properties to reflect student learning. In particular, a set of student mind maps are created three times during a class in both 2011 and 2012 on digital system design using a common technical vocabulary. These mind maps are analyzed by extracting graph metrics by comparison with a criterion mind map, which is an expert created mind map. The metrics are derived from traditional graph metrics (average degree and graph density), three sets of difference metrics analyzed with an internally created tool, and a graph metric invented for comparing proteins. The results of this exploratory analysis is that five of the six metrics can be used to evaluate if a student is learning and connecting the vocabulary in a given subject over time. Additionally, these five metrics are correlated to one another. This result is promising, but we emphasize that these metrics do not correlate directly to class performance based on student grades over the course, and therefore, the current goal for this measurement technique is to be used to provide the student with automated feedback on their mind maps as related to the technical vocabulary of a course.

This work extends our original work by increasing the number of graph metrics that are used to automatically analyze student maps to a criterion map. The idea is to find a number of graph metrics that can then be combined to help analyze a student's mind map and provide them with useful feedback. Even though our results show that compare 5 metrics and each metric can be used to observe student improvement, each of these metrics differs in how the metric can be interpreted and related to the process of learning. Therefore, our goal is to find a number of these metrics so that they can be combined to provide the student with a variety of feedback results to help them understand their errors in terms of the structure of their mind map.

I. INTRODUCTION

One of the limiting aspects of educational research is the techniques available for measuring if a student has learned something. Engineers and their respective research fields, however, are skilled at finding and using measurements to help improve a system. For this reason, it is the author's belief that the community might be best suited for providing more measurement tools to help in improving student learning.

Concept maps [1], [2] and mind maps [3] are useful tools in education, and they can be used both in lectures as classroom assessment techniques (CATs) and outside the classroom for the learner to express ideas in a visual form. Mind maps are

words/concepts connected by lines where a line indicates a relationship between the words, and concept maps have similar structure except the connecting lines tend to have directionality (indicated by arrows) and have associated prepositional phrases that indicate how the words are related. For this work, maps are mathematically called graphs, and therefore, this research can be leveraged in the understanding and characterization of graphs to automatically analyze mind maps.

The two overarching goals are, one, to create tools so that learners can receive automatic feedback on their understanding of a topic, and two, to provide researchers with measurement tools that can be used quantitatively evaluate if an intervention is improving learning beyond existing tools such as tests and surveys. To quantify learning in a map, either the map can be evaluated by an expert, or the map can be compared to a golden map generated by an expert (also known as the criterion map). In this work, the later is used since the first of these evaluation options is very difficult for computers to perform at present. However, this work is a precursor to evaluating open maps.

Over a semester course on digital design, student mind maps were collected at three intervals during the course (course beginning, post exam I, and post exam II). These mind maps were CATs where students had 10 minutes to create mind maps that consisted of 20 technical terms, where these 20 terms are introduced throughout the semester. Each student's mind maps was then digitally encoded and analyzed using various tools. These tools calculate a number of metrics that might reflect learning including:

- A metric based on average degree in a graph
- A metric based on graph density
- A metric based on direct comparison between the two maps for missing and extra connections and words
- A metric of similarity, called Relative Graphlet Frequency distance (RGF-distance)

A more detailed description of these metrics is described including what they are, how they are calculated, and how they might relate to learning within a student's mind map.

The exploratory results show that two of these metrics are highly correlated with each other and that three of these metrics can be used to measure if a student is learning and developing a deeper understanding of the vocabulary over the time of a course. These metrics are checked for correlation

with student grades, but results show that correlation is low and not statistically significant. This is, likely, due to the fact that course grades and deep understanding of the course's vocabulary are not closely related in the chosen course, and a student can perform well in this course independent of vocabulary knowledge.

Some of these ideas were first presented last year at FIE'2012 [4]. This work has been extended to include more data as well as more graph metrics of comparison. The analysis presented in this work shows only the 2012 subjects as the conclusions from the analysis are roughly the same for both 2011 and 2012.

The remainder of this paper is organized as follows. Section II reviews scoring techniques and technical vocabulary. Section III introduces the metrics of measurement in more details and relates how these metrics might be related to student understanding in a mind map. Section IV describes the experiment, and Section V shows results. Finally, Section VI concludes the paper and suggests future work.

II. AUTOMATED MAP SCORING AND TECHNICAL VOCABULARY

As a precursor to automated map analysis, various methods of scoring a map have been proposed by researchers, and a brief list of scoring methods in the literature include the following:

- Concept map - count the number of valid propositions, levels of hierarchy, examples, and crosslinks [1] where weights can be introduced to each count ([5], [6])
- Concept map - a measure of *hierarchiness* which relates hierarchy in the map [7]
- Mind map - compare the scores on tests to the technique [8]
- Concept map - the more important a concept, the closer it is to the top of the tree [9]
- Mind map - have two independent experts score (sometimes with a rubric) the mind map on a scale two times with one week delay and compare correlation of ratings [10]

The scoring of maps has been challenged by many researchers in the literature [11], [12]. Kinchin and Hay [13] criticize the shortcomings of strict scoring of concept maps as a motivation to propose qualitative analysis of maps. Interestingly, one of their key contributions in this qualitative approach is looking for *spoke*, *net*, and *chain* like structures in a map, and these structures are internal graph structures that are captured by what will later be described as graphlets.

Early attempts at automating the analysis of maps and providing feedback focused on hint like mechanisms. A criterion map (called a scaffold in these works) provides student creators with hints for their maps on what is missing and what does not belong [14], [15]. Conlon [16] built a system that used Novak's scoring mechanism and other artificial intelligent concepts to build an open concept map creation system that provided feedback to students. Our goal is study modern graph analysis tools for similar purposes.

A. Technical Vocabulary

In our previous work we reviewed what mind maps and concept maps were and how these maps can be classified and evaluated. For the sake of space we guide the reader to our previous paper [4].

A body of work has focused on the vocabulary and its importance to reading and learning. In this work, there is no attempt to cover this literature, but we provide a short discussion on technical vocabulary. Nation [17] and Coxhead [18] report that the technical vocabulary in a field consists of approximately 5% of the words used in related publications. Chung *et. al.* [19] elaborate on the calculation of this number. Technical vocabulary is more specific than academic vocabulary, and a number of researchers have investigated the impact of specificity and academic vocab on student learning [20], [21], [22], [23].

Though there is a connection between concept and mind maps to vocabulary, there is a limited number of research papers that link or research the combination in any major detail. Beyer *et. al.* [24] describe their work of using concept maps to help new teachers learn and understand the technical vocabulary of education.

III. GRAPH ANALYSIS AND RELATION TO LEARNING

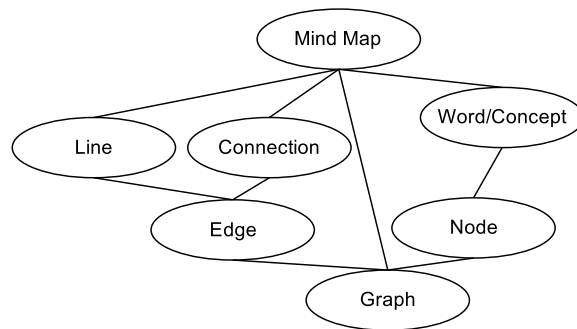


Fig. 1. Example of a mind map on the relationship between mind maps and graphs

Mind maps [3] are a visual tool that can be used to represent the connections between a number of concepts. 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 called the edges of a graph. Mind maps are classified as undirected graphs, which means that the edges do not have a direction, normally indicated by an arrow. In the rest of this section, the graph metrics are discussed including some discussion on why and how each metric captures student understanding. Note that mathematical formulas and definitions are provided in graph notation, but the reader does not need to understand this notation to understand this work.

The first metric is related to *average degree*. The degree of a node is the number of adjacent nodes, or the number of edges that connect to other unique nodes. For example, in figure 1 the degree of the "Mind Map" node is 4 and the degree of the "Connection" node is 2. The average degree is the total

number of edges ($|E|$) divided by the total number of nodes ($|V|$). For this example, the graphs average degree is 9 divided by 7 or, approximately, 1.29. The greater this metric, the more connections on average a node has to other nodes. To make this into a metric of comparison against the criterion map, the following equation is used:

$$AverageDegreeSimilarity = \frac{AverageDegreeStudent}{AverageDegreeCriterion} \quad (1)$$

The *density* of a graph is defined as the total number of edges ($|E|$) divided by the maximum number of edges that a graph could have ($.5 * |V| * (|V| - 1)$). For the graph in figure 1, the maximum number of possible edges is 21 and the number of edges is 9, so the density is 9/21 or, approximately, 0.43. The greater this value, to a maximum of 1, means the more connectivity between the nodes within a graph. A value of 0 means there are no connections, and a value of 1 means the graph is a fully connected graph. This metric of comparison is similar to the previous and follows:

$$DensitySimilarity = \frac{DensityStudent}{DensityCriterion} \quad (2)$$

These first two metrics are measures of how much connectivity there is between the two compared graphs. In terms of relating this to mind maps, the hypothesis is that the more connected a mind map is then the more tightly related the topics are. Comparing student maps and criterion map based on how close the metrics are to one another is an expression of similar global connectivity. The problem, however, is the structure of the mind map is not captured with simple metrics and students might be making wrong connections that somehow make the comparative metrics more equal. Similarly, if a map varies between densely connected areas and sparsely connected areas, then the density and average degree metrics do not capture this and only show the average connectivity.

The third metric is a edge by edge 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), the two graphs can be compared in linear time 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. These statistics are combined in the following equation:

$$GranularSimilarity = \frac{MatchE}{MissN + ExtraE + MatchE} \quad (3)$$

The idea of this equation is it is 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. At present, missing nodes are not included in this equation, but it might be added to the denominator to decrease the similarity factor.

The last metric in this exploration is the RGF-distance, but before defining RGF-distance, graphlets are introduced. Graphlets, formally, are “a connected network with a small number of nodes” [25] and these small graphs are non-isomorphic induced subgraphs of a larger graph. Figure 2

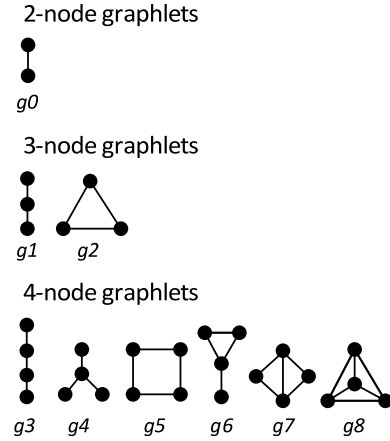


Fig. 2. Graphlets of size 2, 3, and 4

shows all the graphlets of size 2, 3, and 4. Note that the graphlet of size 1 is a single unconnected node and is not that useful.

The power of the graphlet is how it can be used to analyze a graph. The procedure developed by Przulj *et. al.* [25] 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 is constructed in the form ($g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8, \dots, g_{28}, g_{29}$), where g_1 is number of the first type of graphlet of size 3 shown in figure 2 and 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.* [25] and the reader can find the details on the calculation of RGF-distance. The reader should understand that GraphCrunch II will calculate this metric, and the smaller this metric gets means the more similar the two graphs are.

A. Hypotheses on How each Comparison Metric Relates to Learning

Given each of the metrics described above, a set of hypotheses are made to why each metric might capture student understanding in a mind map. Note that it is not expected that any of these metrics can perfectly quantify learning, and the order in which they are presented relates directly to our hypothesis on which metric is, likely, the best choice for measuring learning.

The *AverageDegreeSimilarity* is a connectivity metric that extracts how much connectivity is in the graph on average for both the student and expert generated mind maps. Considering the student mind map, then there are three situations that might emerge as it relates to the criterion map: under-connected, over-connected, and equally connected. In all cases, the average connectivity reflects a student understanding of how the words are related to each other as measured by an average, but this metric does not extract specifics of individual connections. For example, even if this metric results

in “equally connected” (a value of 1), it is possible that the two mind maps are still different, and therefore, this metric only captures a rough estimate of the similarity between student and expert. Note, neither over-connected nor equally connected occurred in the results.

DensitySimilarity is a similar measure of connectivity compared to the first metric. It captures a rough measure of student understanding for the same reasons as described above. In both cases, if these two metrics are considered as feedback tools to a student, then a student might just add or remove edges in their mind map until the metrics equaled one. This is not a true reflection of understanding.

The *GranularSimilarity* metric is created based on an edge to edge comparison between two mind maps, and the breakdown of the metric (as in the instances where an edge is missing or is not needed) would be useful for giving direct student feedback. As a single combined value, the metric reflects a measurement of learning in terms of correctness divided by errors. Both the correct and incorrect edges are valued in terms of one-to-one matching where the measurement reflects how many one-to-one connections have been made correctly. This metric, hypothetically, is better than the previous two, but it does not reflect deeper ideas where the connectivity of a number of words suggests deeper understanding of how these ideas are interrelated.

Finally, *RGF – distance* captures the comparative structure of two graphs, and in terms of comparing student mind maps to the criterion map, the hypothesis is that the lower the RGF-distance means that students better understand the relationships between concepts/words since their maps have more similar structure compared to the expert’s criterion map. Consider the G2 graphlet; this graphlet is a triangle that shows the interconnection between three words, which is a deeper level of understanding then knowing that their is a one-to-one mapping between these words.

IV. SEMESTER MIND MAP EXPERIMENTAL SETUP

For this work, the goal is to analyze student created mind maps over a semester long course and test the respective metrics to see if they quantify how these students are learning the course vocabulary. The focus course for this experiment is a digital design course offered at the 200 level. The course starts with how transistors can be organized to make basic Boolean gates and ends with designing finite state machines using a hardware description language (HDL). From a perspective, the most challenging aspect of the course for most students is the application of HDLs to design hardware as this design language differs significantly from sequential programming languages that students are much more familiar with. However, mind maps only play a small part in understanding how to use HDLs.

These mind map experiments are on closed mind maps. Specifically, a list of 20 technical terms are provided that are introduced in the course. The list consists of 20 terms introduced over the course, but the displayed words are randomized when the students see them.

During the second lecture in the course, mind maps are introduced using an illustration of constructing mind maps

for countries. It is demonstrated how a mind map can be constructed differently depending on if we are treating it as about geographical location, oil supply, or military alliances and enemies. After this basic training, the list of 20 terms is shown and the students have 10 minutes to create their first mind map. This is repeated after exam I and exam II with the same terms and the same amount of time. This means each student who has chosen to participate (IRB approved protocol allowed students to remove their participation agreement any time in the semester before final marks were released) could have created up to three mind maps over the semester.

V. EXPERIMENTAL RESULTS

Three mind map exercises are done over a semester to compare the metrics and analyze how they can measure learning based on three questions. First, do any of these metrics have a strong correlation to grades? Second, is there any correlations between the metrics themselves? Finally, in the last section, how do the metrics change on a per student basis over time as a reflection of learning?

A. Metrics Correlation to Grades

The first hypothesis is that vocabulary understanding measured by the proposed graph metrics will have a correlation with grades.

TABLE I. CORRELATION RESULTS BETWEEN COMPARISON METRICS AND GRADES

Metric	Correlation Values			
	Exam I	Exam II	Exam III	Final Grades
<i>AverageDegreeSimilarity</i>				
Pre	0.04	0.04	0.14	0.22
Post Exam I	0.03	0.04	0.04	0.01
Post Exam II	0.02	0.18	0.04	0.04
<i>DensitySimilarity</i>				
Pre	0.01	0.10	0.05	0.21
Post Exam I	0.02	0.00	0.19	0.06
Post Exam II	0.01	0.12	0.09	0.04
<i>GranularSimilarity</i>				
Pre	0.07	0.23	0.07	0.25
Post Exam I	0.09	0.18	0.06	0.11
Post Exam II	0.05	0.09	0.07	0.00
<i>RGF – distance</i>				
Pre	0.16	0.12	0.09	0.20
Post Exam I	0.19	0.13	0.01	0.01
Post Exam II	0.11	0.19	0.13	0.06

Table I summarizes the results of the correlation data between comparison metrics and grades. The first column contains the labels for each metric calculated with the mind maps from Pre, post Exam I, post Exam II. Columns two through five contain the correlation coefficient calculated between the comparison metric and the students grade on Exam I, II, III, and their final grade. The bold values are statistically significant for $p < 0.01$ where the populations are $n = 41$ for Pre, $n = 39$ for post Exam I, and $n = 36$ for post Exam II (two-tailed test).

From these results, there are no statistically significant points. In general, we state that there does not appear to be a strong correlation between the comparison metrics and grades. These results suggest there might be some deeper connection that might need more investigation if the technical terminology is directly related to course grades.

Since the course studied does not have a strong dependence on technical vocabulary and is graded based on the performance of the students in designing and solving problems, we don't think there should be a strong correlation. However, we are, presently, looking at other courses where there is a stronger connection between grades and understanding of technical language to see if there is stronger correlation between our metrics and grades.

B. Correlation between Comparison Metrics

In this section, the question is how do the different measurement metrics relate to one another. Earlier discussion described how average degree and density are similar graph metrics in terms of connectivity. Here, any close relationships between all the metrics is explored.

Table II shows the correlation of each of the four metrics to each other as compared based on the same mind map creation time. Column 1 contains the time when the mind map was created, and columns 2 and 3 contain the label of the measurement metric being compared. Finally, the last column contains the correlation value where bold values indicate the correlation is statistically significant for $p < 0.01$. The population values are the same as the previous analysis.

The results show that there are a number of statistically significant correlations, but to consider that two metrics are closely related to one another we would expect correlation over all three time points. This happens in the three way relationship between *AverageDegreeSimilarity*, *GranularSimilarity*, and *RGF - distance*, which suggests that these three metrics are capturing similar results.

C. Per Student Metric Analysis

Since there is no strong correlation between grades and our comparison metrics, quantifying learning might be viewed from the perspective of an improvement in individual performance. In this analysis section, we plot the metrics for each student's map over time with the hypothesis that the metric will change for the better as the student progresses and learns in a course. Since there is a significant amount of data for this analysis, we will only show students who finished our course with a grade between 80% and 90% who did all three mind maps ($n = 15$ students). These results have similar behavior for all students in the class and demonstrates some interesting trends.

Before looking at the results we must make an assumption to interpret them. The assumption is that students will, on average, create mind maps that are more similar to the criterion map as the course progresses. This assumption is debatable, but from our qualitative experiences with student mind maps we believe this assumption is true.

Figure 3 shows the four graphs, one for each comparison metric, plotting a line that represents a student's performance over time related to the respective metric. In each case, the x-axis is time in terms of when the mind map was created (Pre, Post Exam I, and Post Exam II), and the y-axis is a scale for the respective metric. The values on this scale are not very important, but we briefly describe how the change of a metric reflects improvement relative to the criterion map. In

other words, how should the metric change to show that the student mind map is more similar to the criterion map. For *AverageDegreeSimilarity* and *DensitySimilarity* plots (upper left and lower left in Figure 3) an improvement in understanding happens when the metric approaches 1, but the value can be greater or less than 1 since the metric is a ratio between student and criterion maps. For *GranularSimilarity* plot (upper right in Figure 3), an improvement in understanding happens as the metric approaches 1 (where the value can only be less than 1). Finally, for *RGF - distance* plot (lower right in Figure 3) an improvement in understanding happens as the metric decreases in value.

Of the four metrics, the two plots for *GranularSimilarity*, and *RGF - distance* show that on average these students are learning as the students respective metric measurements show improvement. This can be observed quite clearly for *GranularSimilarity*.

AverageDegreeSimilarity and *DensitySimilarity* do not follow this clear pattern of improvement, and from the results, we conclude that this metric do not offer an easy visualization of learning. For the graph *DensitySimilarity*, it appears that the trends are random, and we can conclude that this metric, as yet, can not capture learning. On the other hand, *AverageDegreeSimilarity* has some trends (and as previously shown, correlates to the good metrics), and therefore, this metric should be further studied to be used as one way of reflecting improvement.

Another interesting observation from the two most appropriate graphs for visualizing learning (*GranularSimilarity*, and *RGF - distance*) is the shape of a student's line. In some cases, the student has made a more similar mind map compared to the criterion map after exam I then to after exam II, and this might be phrased as the student seems to have lost some of their knowledge. For example, student 12_11 and 12_37 show this behavior. Their are a number of hypotheses that could be made. For example, the mind maps created after exam II included more vocabulary, which the student was unsure how to connect with previous vocabulary. This is a question for future work.

VI. DISCUSSION AND CONCLUSION

The last analysis shows that three comparison metrics can show that a student is learning the technical vocabulary from automated analysis. However, as discussed in earlier sections we believe that both the *GranularSimilarity* and *RGF - distance* metrics are the most useful in terms of giving students automatic feedback about the quality of their mind maps and as a visualization tool it seems clear that *GranularSimilarity* may be the best visualization tool.

From the perspective of using this tool in educational research, at present the three metrics can be used to provide a true/false answer on whether a student's understanding of concepts as reflected by their mind maps has improved. As of yet, there is no quantity that can be associated with this improvement as in percentage improvement. In the future, as we increase the data sets and improve these metrics, we may be able to provide such quantities. At present, though, we believe that these techniques should be used to validate if students are improving, but these techniques should neither be used for

TABLE II. CORRELATION BETWEEN METRICS

Time	Metric 1	Metric 2	Correlation
Pre	<i>AverageDegreeSimilarity</i>	<i>DensitySimilarity</i>	0.27
Post Exam I	<i>AverageDegreeSimilarity</i>	<i>DensitySimilarity</i>	0.86
Post Exam II	<i>AverageDegreeSimilarity</i>	<i>DensitySimilarity</i>	0.95
Pre	<i>AverageDegreeSimilarity</i>	<i>GranularSimilarity</i>	0.58
Post Exam I	<i>AverageDegreeSimilarity</i>	<i>GranularSimilarity</i>	0.70
Post Exam II	<i>AverageDegreeSimilarity</i>	<i>GranularSimilarity</i>	0.76
Pre	<i>AverageDegreeSimilarity</i>	<i>RGF – distance</i>	0.82
Post Exam I	<i>AverageDegreeSimilarity</i>	<i>RGF – distance</i>	0.84
Post Exam II	<i>AverageDegreeSimilarity</i>	<i>RGF – distance</i>	0.83
Pre	<i>DensitySimilarity</i>	<i>GranularSimilarity</i>	0.07
Post Exam I	<i>DensitySimilarity</i>	<i>GranularSimilarity</i>	0.56
Post Exam II	<i>DensitySimilarity</i>	<i>GranularSimilarity</i>	0.73
Pre	<i>DensitySimilarity</i>	<i>RGF – distance</i>	0.11
Post Exam I	<i>DensitySimilarity</i>	<i>RGF – distance</i>	0.66
Post Exam II	<i>DensitySimilarity</i>	<i>RGF – distance</i>	0.83
Pre	<i>GranularSimilarity</i>	<i>RGF – distance</i>	0.59
Post Exam I	<i>GranularSimilarity</i>	<i>RGF – distance</i>	0.64
Post Exam II	<i>GranularSimilarity</i>	<i>RGF – distance</i>	0.63

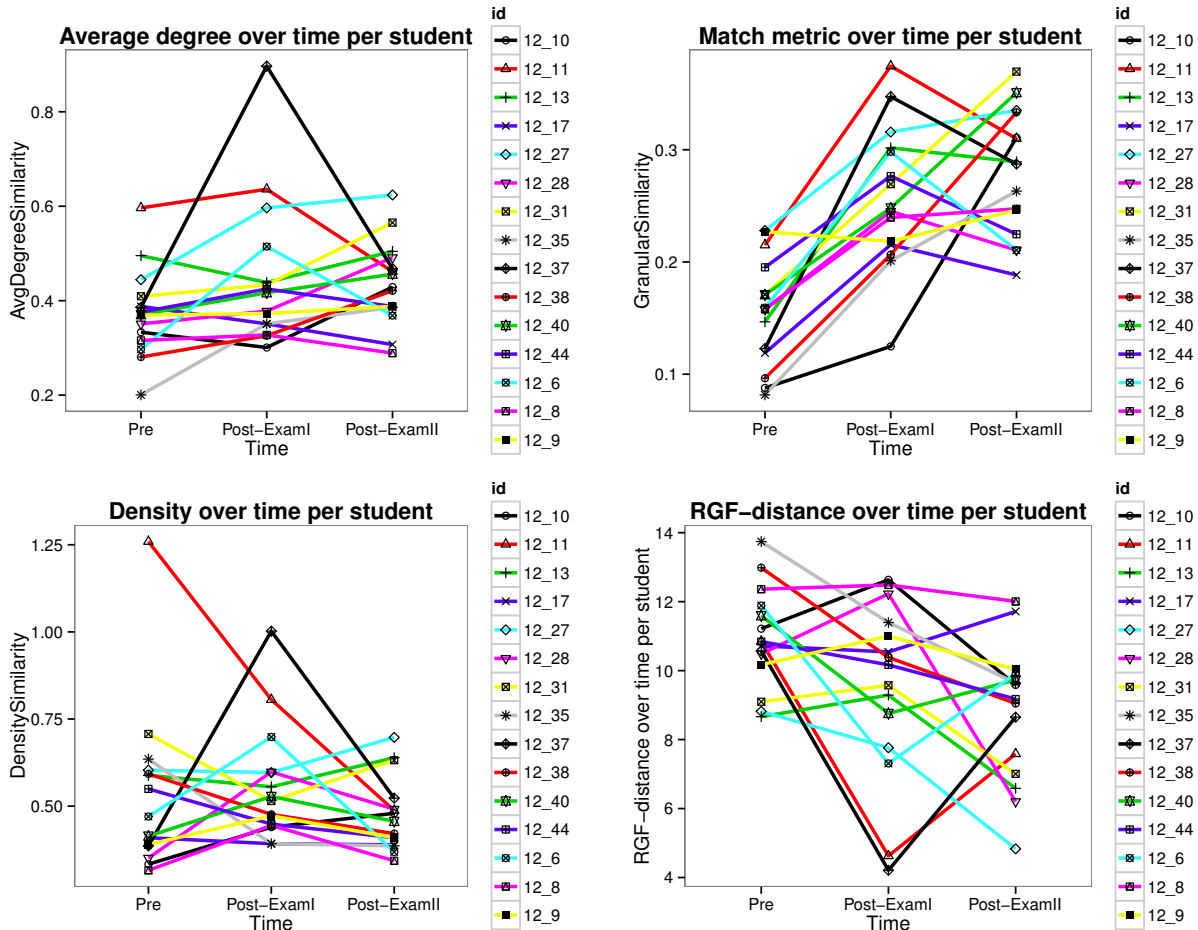


Fig. 3. Line graphs for student mind maps RGF-distance over semester for each comparison metric.

formal assessment of grades nor a quantitative value of how much has been learned.

A. Conclusion

In this work, an exploration of the value of four comparison metrics was studied that could be used to automatically analyze student’s mind maps compared to an expert’s criterion mind

map. Each of these metrics was described including how they might be related to learning. Student mind maps were analyzed from a class to understand if these metrics provide any useful measurements.

The results show that there is no strong correlation between grades and any of the metrics. Additionally, there is some relation between all of these metrics as shown by correlations between them. Most importantly, it was observed how these metrics change over time on a per student basis. These results suggest that three of these metrics, *Average Degree Similarity*, *Granular Similarity*, and *RGF – distance* can be used to determine if a student is improving their understanding of the relationships between vocabulary.

For future work, we are extending this experiment to other disciplines and their respective courses as we can find interested faculty members. In this process and with our new matching metric, we have learned that one of the key issues for further research is how to select the technical vocabulary to include in an experiment. Our plan is to further investigate this question and learn to provide a good methodology for teachers to select their vocabulary among other additional questions identified in the paper.

REFERENCES

- [1] J. D. Novak and D. B. Gowin, *Learning how to learn*. New York: Cambridge University Press, 1984.
- [2] D. F. Dansereau and C. D. Holley, *Development and Evaluation of a Text Mapping Strategy*. Amsterdam, New York & Oxford: North-Holland Publishing Company, 1982, pp. 536–554.
- [3] T. Anderson, *Study skills and learning strategies*. Center for the Study of Reading, University of Illinois at Urbana-Champaign, 1978. [Online]. Available: <http://books.google.com/books?id=JNHUjwEACAAJ>
- [4] P. Jamieson, “Using modern graph analysis techniques on mind maps to help quantify learning,” in *Frontiers in Education Conference (FIE), 2012*, oct. 2012. [Online]. Available: http://www.users.muohio.edu/jamiespa/html_papers/fie_12.pdf
- [5] T. Stoddart, R. Abrams, E. Gasper, and D. Canaday, “Concept maps as assessment in science inquiry learning - a report of methodology,” *International Journal of Science Education*, vol. 22, no. 12, pp. 1221–1246, 2000. [Online]. Available: <http://www.informaworld.com/openurl?genre=article&doi=10.1080/095006900750036235&magic=crossref>
- [6] D. C. West, J. K. Park, J. R. Pomeroy, and J. Sandoval, “Concept mapping assessment in medical education: a comparison of two scoring systems,” *Medical Education*, vol. 36, no. 9, pp. 820–826, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12354244>
- [7] 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,” University of California, Tech. Rep., 1997. [Online]. Available: <http://research.cse.ucla.edu/Reports/TECH455.PDF>
- [8] I. Abi-El-Mona and F. Adb-El-Khalick, “The influence of mind mapping on eighth graders science achievement,” *School Science and Mathematics*, vol. 108, no. 7, pp. 298–312, 2008. [Online]. Available: <http://dx.doi.org/10.1111/j.1949-8594.2008.tb17843.x>
- [9] D. Leake, A. Maguitman, and T. Reichherzer, “Understanding knowledge models: Modeling assessment of concept importance in concept maps,” in *In Proceedings of CogSci 2004. Mahwah, NJ: Erlbaum. In*, 2004, pp. 785–800.
- [10] E. Evrekli, D. Inel, and A. Galim, “Development of a scoring system to assess mind maps,” *Procedia - Social and Behavioral Sciences*, vol. 2, no. 2, pp. 2330 – 2334, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S187704281000371X>
- [11] M. A. Ruiz-Primo and R. J. Shavelson, “Problems and issues in the use of concept maps in science assessment,” *Journal of Research in Science Teaching*, vol. 33, no. 6, pp. 569–600, Aug. 1996. [Online]. Available: [http://dx.doi.org/10.1002/\(SICI\)1098-2736\(199608\)33:6%3C569::AID-TEA1%3E3.0.CO;2-M](http://dx.doi.org/10.1002/(SICI)1098-2736(199608)33:6%3C569::AID-TEA1%3E3.0.CO;2-M)
- [12] X. Liu and M. Hinchey, “The internal consistency of a concept mapping scoring scheme and its effect on prediction validity,” *International Journal of Science Education*, vol. 18, no. 8, pp. 921–937, 1996. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/0950069960180805>
- [13] 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*, vol. 42, pp. 43–57, 2001. [Online]. Available: <http://www.personal.psu.edu/kmo178/blogs/kmorourke/qualitative%20approach%20to%20concept%20map%20analysis.pdf>
- [14] K. Chang, Y. Sung, and S. Chen, “Learning through computer-based concept mapping with scaffolding aid,” *Journal of Computer Assisted Learning*, vol. 17, no. 1, pp. 21–33, 2001. [Online]. Available: <http://dx.doi.org/10.1111/j.1365-2729.2001.00156.x>
- [15] G. Biswas, D. Schwartz, and J. Bransford, “Smart machines in education,” K. D. Forbus and P. J. Feltovich, Eds. Cambridge, MA, USA: MIT Press, 2001, ch. Technology support for complex problem solving: from SAD environments to AI, pp. 71–97. [Online]. Available: <http://www.vuse.vanderbilt.edu/~biswas/Research/ile/papers/sad01/sad01.pdf>
- [16] T. Conlon, “But is our concept map any good?: Classroom experiences with the reasonable fallible analyser,” in *Proc. of the First Int. Conference on Concept Mapping*, 2004. [Online]. Available: <http://cmc.ihmc.us/papers/cmc2004-054.pdf>
- [17] I. Nation, *Learning Vocabulary in Another Language*. Cambridge University Press, 2002. [Online]. Available: <http://books.google.com/books?id=rg8Aue4jJWkC>
- [18] A. Coxhead, *Essentials of teaching academic vocabulary*. Houghton Mifflin Company, 2006.
- [19] T. Chung and P. Nation, “Technical vocabulary in specialised texts,” *Reading in a foreign language*, vol. 15, no. 2, pp. 103–116, 2003.
- [20] K. HYLAND and P. TSE, “Is there an academic vocabulary?” *TESOL Quarterly*, vol. 41, no. 2, pp. 235–253, 2007. [Online]. Available: <http://dx.doi.org/10.1002/j.1545-7249.2007.tb00058.x>
- [21] K. Hyland, “Specificity revisited: how far should we go now?” *English for Specific Purposes*, vol. 21, no. 4, pp. 385 – 395, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S088949060100028X>
- [22] M. Lessard-Clouston, “Strategies and success in technical vocabulary learning: Students’ approaches in one academic context,” *Indian Journal of Applied Linguistics*, vol. 34, no. 1/2, p. 31, 2008.
- [23] J. M. Harmon, W. B. Hedrick, and K. D. Wood, “Research on vocabulary instruction in the content areas: Implications for struggling readers,” *Reading & Writing Quarterly*, vol. 21, no. 3, pp. 261–280, 2005. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10573560590949377>
- [24] B. A. Beyerbach, “Developing a technical vocabulary on teacher planning: Preservice teachers’ concept maps,” *Teaching and Teacher Education*, vol. 4, no. 4, pp. 339 – 347, 1988. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0742051X88900327>
- [25] N. Przulj, D. G. Corneil, and I. Jurisica, “Modeling interactome: scale-free or geometric?” *Bioinformatics*, vol. 20, no. 18, pp. 3508–3515, 2004.