

Deriving Individual and Group Knowledge Structure from Network Graphs and from Essays

Roy B. Clariana

The Pennsylvania State University, Instructional Systems Program

RClariana@psu.edu

Abstract This chapter describes an approach for deriving an individual's knowledge structure as a data association array derived from texts such as essays or from concept maps. We describe two free software tools that were designed to work with Pathfinder Network analysis, *ALA-Mapper* which processes concept maps (network graphs) and *ALA-Reader* which processes text into proximity arrays. These arrays can then be combined into a group representation, for example to compare several novices to an expert, or all males to all females in a course section, or management to employees. Several investigations are reviewed that have used these tools. Based on this, other possible applications and also limitations of the software and the approach are described. Finally we describe the next stage of software development of these tools.

Keywords: concept maps, network graphs, knowledge structure

Introduction

This volume is about computer-based approaches to systematically measure knowledge and especially complex knowledge that is less open to traditional test and measurement approaches. This chapter focuses on two apparently fundamentally related ways to elicit what people know, network graphs (usually as concept maps) and essays. Our ideas originated as a method to score concept maps. Because of the visual similarity between concept maps and Pathfinder networks (*PFNETs*), the Pathfinder Network scaling tool set, approach, and research base and its underlying premises from graph theory were used to develop software (e.g., *ALA-Mapper*, 2002) to score participants' concept maps relative to an expert's map. Because concept maps are frequently used in classrooms to replace outlining as an organizational aid for writing essays, we became interested in measuring the relationship between concept maps and the essays derived from these maps, and it was a natural progression to develop software based on our concept map scoring approach to score essays (*ALA-Reader*, 2004). This chapter begins by describing the *ALA-Mapper* and *ALA-Reader* scoring approach, then the investigations with these two tools are reviewed, and finally suggestions for future research are provided.

Pathfinder Network Analysis

Pathfinder network analysis is a well established system for representing the organization of knowledge that has been applied in many domains (Jonassen, Beissner, & Yacci, 1993; Schvaneveldt, 1990). Pathfinder algorithms converts estimates of relatedness of pairs of terms into a network representation of those terms called Pathfinder network graphs (*PFNETs*) that are usually 2-dimensional representations of a matrix of relationship data in which concept terms are represented as nodes (also called vertices) and relationships are shown as weighted links connecting the nodes (links are also called edges). *PFNETs* closely resemble concept maps in appearance, but without link labels. The set of links derived by the analysis is determined by the patterns in the raw proximity data and these are influenced by two algorithm control parameters, q and Minkowski's r .

There are three steps in the Pathfinder approach. In Step 1, raw proximity data is collected typically using a word-relatedness judgment task. Participants are shown a set of terms two at a time, and judge the relatedness of each pair of terms for example on a scale from one (low) to nine (high). The number of pair-wise comparisons that participants must make is $(n^2 - n)/2$, with n equal to the total number of terms in the list.

In Step 2, Knowledge Network and Orientation Tool for the Personal Computer (*KNOT*, 1998) is used to reduce the raw proximity data into a *PFNET* representation. The Pathfinder algorithm determines a least-weighted path that links all of the terms. The rules for calculating the least-weighted path can be adapted by adjusting parameters that reduce or prune the number of links in the resulting *PFNET* (refer to Dearholt & Schvaneveldt, 1990). The resulting *PFNET* is based on a data reduction approach that is purported to represent the most salient relationships in the raw proximity data.

In Step 3, the comparison of the participant's *PFNET* to an expert or alternative referent *PFNET* is calculated also using *KNOT* software (Goldsmith & Davenport, 1990). The two most commonly reported similarity measures are Common and Configural Similarity. Common is the sum of the links shared by two *PFNETs* (the intersection of two *PFNETs*). Configural Similarity, which is also called neighbourhood similarity, is obtained by averaging the average of the intersection divided by the union for every term in the two *PFNETs*. Usually one of the *PFNETs* is an expert referent and the other is the participant's *PFNETs*, thus Common and Configural Similarity show the relationship between the participant's concept map and the expert's concept map.

Note that *KNOT* has a group average feature that averages multiple proximity files to obtain a group average *PFNET* representation. This feature is especially useful for comparing one group to another or for comparing different groups to some referent. Our experience is that group average *PFNETs* are more robust than individual *PFNETs*. Averaging seems to remove idiosyncratic and error responses in individual *PFNETs*.

ALA-Mapper

What information components of concept maps can be collected automatically by a computer and how can the Pathfinder approach be used to score concept maps? Concept maps consist of terms, links, and link labels (i.e., propositions); also there is an overall visual layout of the map that consists of patterns of links and also closeness of terms. *ALA-Mapper* uses either links between terms or distances between terms as an alternative to word-relatedness judgment tasks in Step 1 for obtaining raw proximity data, while Steps 2 and 3 are conducted in the conventional way. Thus, the main contribution of *ALA-Mapper* is its capability to convert components of a concept map into raw proximity data in Step 1 of the Pathfinder analysis approach.

So what information in concept maps can be measured? There are at least three or four different cognitive tasks involved when creating a concept map that may leave a residue in the map. First, if the map is ‘open ended’ where students may use any terms in their map, then a critical task is recalling (or possibly recognizing from a list) the most important terms/concepts to include in their map. Alternately, if a list of terms is provided and the students are told to use all of the terms (fixed or closed mapping), then recall of terms is not a factor. Note that it is easier for both instructors and computers to score closed maps compared to open maps. Next, students must group related terms together, often in an intuitive way, and this most likely relates to their internal network structure. Then students identify propositions by linking pairs of terms with a line and adding a linking phrase, that reflect the meaning of the proposition in that context. While students work on the later stages of their map, they continually revise small components of their map making it easier to grasp, and this also seems to be an intuitive activity of making it ‘feel’ right and reflects both the structure of their knowledge and an internalized graphic grammar or norm of what things like this should look like (Clariana & Taricani, *in press*).

Probably the basic meaningful psychological components of concept maps (and essays) are propositions (Einstein et al., 1984; Kintsch, 1974; Kintsch and van Dijk, 1978). In essence, a proposition consists of a subject-(verb)-object which in a concept map consists of a term-(linking phrase)-term. However, link labels are more computationally difficult to collect, handle, compare, and analyze than just the presence or absence of a link between terms, so *ALA-Mapper* pragmatically uses links rather than link labels. Note that linking phrases may not be critical for concept map analysis. Harper, Hoef, Evans, and Jentsch (2004) reported that the correlation between just counting link lines (i.e., node–node) compared to counting valid propositions (i.e., node–label–node) in the same set of maps was $r = 0.97$, suggesting that link labels add little additional information over just counting links. *ALA-Mapper* converts node–node information in concept maps and other types of network graphs into two separate kinds of raw data, links between terms (note: without link labels) and distances between terms measured in screen pixels (see Figure 1). This raw data can then be analyzed by Pathfinder analysis.

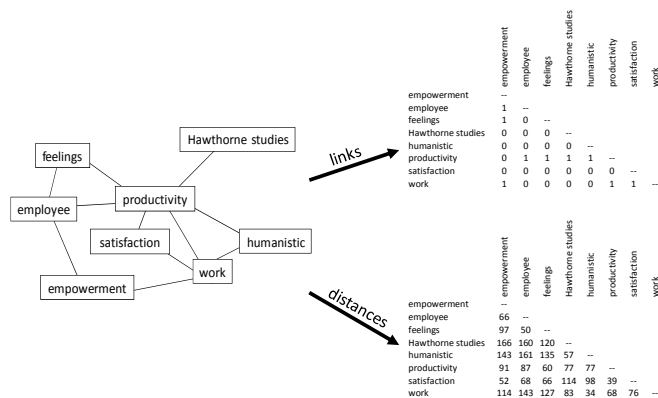


Figure 1. A sample concept map like network graph converted into a link array and a distance array.

Clariana, Koul, and Salehi (2006) used *ALA-Mapper* for scoring open-ended concept maps. Practicing teachers enrolled in graduate courses constructed concept maps on paper while researching the topic, “the structure and function of the heart and circulatory system” online. Participants were given the online addresses of five articles that ranged in length from 1,000 to 2,400 words but were encouraged to view additional resources. After com-

pleting their research, participants then used their concept map as an outline to write a 250-word text summary of this topic. *ALA-Mapper* was used to measure the geometric distances between terms in the concept maps, referred to as distance data, and to count the link lines that connected terms, referred to as link data (see Figure 1). Using Pathfinder *KNOT* software, the raw distance and link data were converted into *PFNETs* and then were compared to an expert's *PFNETs* to obtain similarity scores. Five pairs of raters using rubrics also scored all of the concept maps and text summaries. The Pearson correlation values for the concept maps scored by raters compared to: (a) concept map link-based scores were 0.36, (b) concept map distance-based scores were 0.54, and (c) text summaries scored by raters was 0.49; thus the *ALA-Mapper* distance scores were a bit more like the raters' concept map scores than were the link scores. The correlation values for the text summaries scored by raters compared to (a) concept map link-based scores were 0.76 and (b) concept map distance-based scores were 0.71; thus the *ALA-Mapper* link and distance scores were both quite like the raters' text summary scores.

In a follow-up study, Poindexter and Clariana (2004) used the same Pathfinder scoring technique but asked participants to draw network maps rather than concept maps. Network maps are like concept maps, except that there are no linking phrases. The mapping directions specifically directed the participants to use spatial closeness to show relationships and intentionally deemphasized the use of link lines. Participants completed one of three print-based text lessons on the heart and circulatory system. The three lesson treatments included adjunct constructed response questions (an item specific approach that emphasizes propositions), scrambled-sentences (a relational approach that emphasizes concept associations), and a reading only control. Participants completed three multiple-choice posttests that assessed identification, terminology, and comprehension and then were handed a list of 25 pre-selected terms and were asked to draw a network map. The adjunct question treatment was significantly more effective than the other lesson treatments for the comprehension outcome, and no other treatment comparisons were significant. *ALA-Mapper* scores based on link data were more related to terminology ($r = .77$) than to comprehension ($r = .53$), while scores based on distance data were slightly more related to comprehension ($r = .71$) than to terminology ($r = .69$). The links drawn to connect terms related to verbatim knowledge from the lesson text covering facts, terminology, and definitions, while the distances between terms related to the broader processes and functions of the heart and circulatory system.

In a follow-up study (Taricani & Clariana, 2006), sixty undergraduate students read a print-based instructional text on the heart and circulatory system and then created concept maps of that content. Half were given feedback in the form of a prepared hierarchical concept map and the other half, no feedback map. Then all completed a multiple-choice posttest with 20 terminology and 20 comprehension questions. The concept maps were scored using *ALA-Mapper* and these concept map scores were compared to the terminology and comprehension posttest scores. Similar to Poindexter and Clariana (2004) above, concept map scores derived from link data were more related to terminology ($r = .78$) than to comprehension ($r = .54$) whereas concept map scores derived from distance data were more related to comprehension ($r = .61$) than to terminology ($r = .48$).

Several other investigations in this volume (see Chapter x) have also used *ALA-Mapper* type distance data with Pathfinder analysis for various purposes, for example, to measure the effects of lesson screen background colour and of generating lesson headings on posttest lesson topic structure. The idea is that there is worthwhile relational information in the distances between terms in a network graph. However, our view is that this distance information is fragile, and that pre-map training or strong directions that emphasize proposition-specific elements in the maps damages the distance information. For example, concept map training that demands that all map elements be propositions that are term-(linking phrase)-term turn the participants' focus to those elements and away from distance-related relational aspects of their knowledge. Ironically, if the rubrics used to score these concept maps are also strongly proposition oriented, then those maps that do have a focus on propositions will score relatively higher, thus confirming that propositions are key elements in concept maps. In contrast, Poindexter and

Clariana (2004) described above provided mapping directions that intentionally de-emphasized propositions (links and linking phrases were optional) and emphasized distances between terms, with the result that the distance scores obtained larger correlations with the comprehension measures. Thus investigators must be sensitive to the relational or proposition specific effects of their pre-map training, their mapping directions, and the rubrics used to score the maps.

Rubrics and Network Graph Scores

So far, the investigations above have avoided the issue of what precisely are the raters scoring when they score a concept map or other type of network graph. These studies used holistic scoring that only considered the content accuracy reflected in the concept maps, a quantitative approach where more correct ideas usually obtains a higher score from the raters. However, Koul, Clariana, and Salehi (2006) reported that *ALA-Mapper* data correlated better with raters' scores (note: 11 pairs of raters) using a qualitative than a quantitative rubric. Teachers enrolled in a graduate course working in pairs researched a science topic online and created a concept map of the topic. Later, participants individually wrote a short essay from their concept map. The concept maps and essays were scored by the computer based tools and by human raters using qualitative and quantitative rubrics. The quantitative rubric was adapted from the Lomask et al. (1992) rubric. This rubric considered *size* (the count of terms in a student map expressed as a proportion of the terms in an expert concept map) and *strength* (the count of links in a student map as a proportion of necessary, accurate connections with respect to those in an expert map). The qualitative rubric for scoring concept maps was based on research by Kinchin and Hay (2000). This rubric deals with three common map structure which may be interpreted as indicators of progressive levels of understanding: 1) *Spoke*, a structure in which all the related aspects of the topic are linked directly to the core concept, but are not directly linked to each other; 2) *Chain*, a linear sequence of understanding in which each concept is only linked to those immediately above and below; and 3) *Net*, a network both highly integrated and hierarchical, demonstrating a deep understanding of the topic. Computer-based concept map scores were a very good measure of the qualitative aspects of the concept maps (link $r = 0.84$ and distance $r = 0.53$) and were an adequate measure of the quantitative aspects (link $r = 0.65$ and distance $r = 0.50$).

These various results were evidence to convince us that *ALA-Mapper* scores were not really concept map 'content' scores, but rather that *ALA-Mapper* scores are a measure of structural knowledge that correlates somewhat with some forms of concept map content scores as well as with different kinds of traditional posttests. Also, the distance data and the link data in network graphs both contain interesting and useful information. However, the internal and external context factors can enhance or suppress the information content in this raw data and so must be well controlled.

ALA-Reader

The *ALA-Reader* essay analysis approach was adopted directly from the *ALA-Mapper* concept map analysis approach. *ALA-Reader* searches for key terms in text and forms links in an aggregate array (a) between all terms that occur in the same *sentence* or else (b) between consecutive terms in a *linear* pass through the text. The aggregate array raw data for a text processed by *ALA-Reader* is similar in form to the network graph link raw data (see the top right portion of Figure 1) and is analyzed by Pathfinder *KNOT* software in the same way.

Compare-contrast type essay questions have been used to assess relational understanding that is part of knowledge structure (Gonzalvo, Canas, & Bajo, 1994) though any text likely is influenced by the author's knowledge structure. Goldsmith, Johnson, and Acton (1991) state "Essay questions, which ask students to discuss the

relationships between concepts, are perhaps the most conventional way of assessing the configural aspect of knowledge.” (p.88) It is rather critical to keep in mind that an essay contains different kinds of information, and that the scoring approach determines what is actually measured and most if not all essay scoring approaches, human or computer-based, do not intentionally measure knowledge structure. But whether it is intentionally measured or not, essays contain at least a reflection of an individual’s knowledge structure.

Sentence Aggregate Approach

The *ALA-Reader sentence aggregate approach* was developed to analyze at the sentence level because sentences are an important unit of text organization. Sentences contain one or more propositions and the sentence aggregation approach seeks to capture the important node–node associations represented by propositions in sentences. To analyze sentences in text, first *ALA-Reader* disregards all of the words in the text except for preselected key terms, and replaces the synonyms and metonyms of key terms with its appropriate key term. Then the key terms that co-occur in the same sentence are represented in a proximity array, the lower triangle of an n-by-n array containing $n(n-1)/2$ elements. A “1” entered in the proximity array indicates that two key terms co-occurred in the same sentence and a “0” means those two key terms did not occur in the same sentence. The software continues to aggregate sentences into the proximity array until all of the text is processed. For example, given four sentences of a participant’s essay regarding the humanistic management approach from Clariana, Wallace, and Godshalk (2008) shown with key terms or their synonyms underlined,

Humanists believed that job satisfaction was related to productivity. The Hawthorne studies tried to determine if lighting caused people to be more productive employees. However, it was found that employees valued being selected to participate in the study and were more productive when they felt “special.” They found that if employees were given more freedom and power in their jobs, then they produced more.

These four sentences would be reduced by *ALA-Reader* to this sentence aggregate array of selected key terms shown in Figure 2. The force directed network graph of the four sentence aggregate is shown in Figure 3.

	<u>Sentence 1</u>								<u>Sentence 1 & 2</u>								<u>Sentence 1, 2, & 3</u>								<u>Sentence 1, 2, 3, & 4</u>							
	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work
empowerment	1								1								1								1							
employee	0	1							0	1							0	1							0	1						
feelings	0	0	1						0	0	1						0	1	0						0	1	0					
Hawthorne studies	0	0	0	1					0	1	0	1					0	1	0	1					0	1	0	1				
humanistic	0	0	0	0	1				0	0	0	0	1				0	0	0	0	1				0	0	0	0	1			
productivity	0	0	0	0	1	1			0	1	0	1	1	1			0	1	1	1	1	1			1	1	1	1	1			
satisfaction	0	0	0	0	1	1	1		0	0	0	0	1	1	1		0	0	0	0	1	1	1		0	0	0	0	1	1		
work	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	

Figure 2. The sentence aggregate proximity file after each sentence is processed by *ALA-Reader*.

Sentence 1, 2, 3, & 4

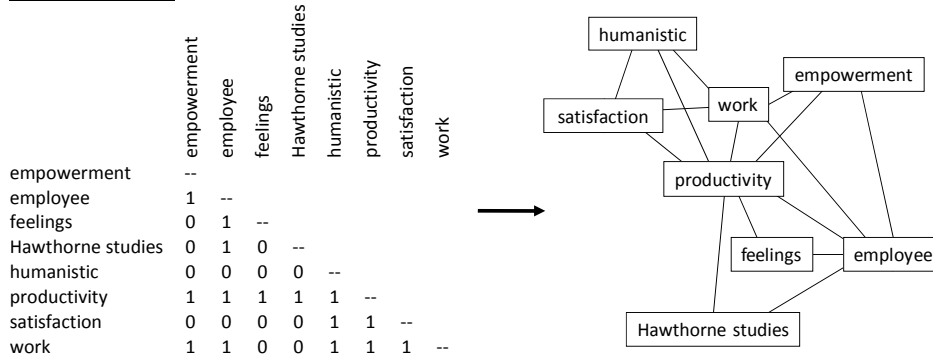


Figure 3. The force directed network graph of the four sentence aggregate.

Linear Aggregate Approach

The *ALA-Reader linear aggregation approach* sets links (1s) in the aggregation array between adjacent terms during a linear pass through the text and so will always obtain a connected graph, but the result is almost certainly not just a linear chain of words. As important words are used multiple times in the essay passage, those terms will have more links coming in and going out, and the structure when represented as a force-directed network graph begins to fold and draw related terms closer together in the 2-dimensional space. The same text used above in Figure 2 would be reduced by *ALA-Reader* to this linear sequence of selected key terms (with link numerical order): “humanistic -1- work -2- satisfaction -3- productivity -4- Hawthorne studies -5- productivity -6- employee -7- employee -8- productivity -9- feelings -10- employee -11- empowerment -12- work -13- productivity” (see the linear visual representation of this sequence in Figure 4).

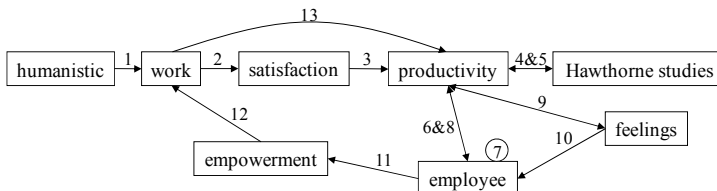


Figure 4. A visual representation of the linear sequence of key terms from a participant’s essay passage about the Humanistic management approach.

The linear sequence of terms can also be represented as a force-directed network graph (a *PFNET*) that suggests the more and less salient relationships in the passage based on the degree of the nodes but also provides some idea of possible indirect relationships based on spatial closeness (see Figure 5). For example, the term “productivity” with 6 links is a high degree node (i.e., with 3 or more links) and so is a central node in this *PFNET*. This indicates that the student’s essay passage describes the humanistic management approach in terms of its relationship to productivity. The terms “work” and “employee” are also high degree nodes and so are also important terms. “work” is a high degree node with 6 links and so is a central node in this *PFNET*.

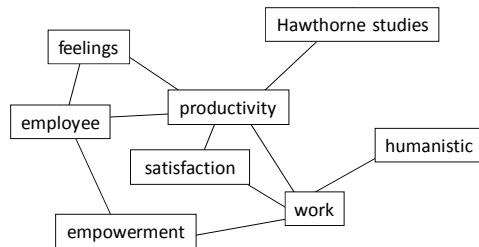


Figure 5. The force-directed network graph (*PFNET*) of linear representation of the participant's essay passage shown in Figure 4.

Clariana and Koul (2004) used *ALA-Reader* software to score students' essays on the structure and function of the heart and circulatory system relative to an expert's essay. At that time the software could only analyze using the sentence aggregate approach. For benchmark comparison, the essays were also scored by 11 pairs of human raters and these 11 scores were averaged together into one composite essay score. Compared to the human raters' composite score, the *ALA-Reader* scores were 5th out of 12, with an $r = .69$ (the 12 scores correlations ranged from $r = .11$ to $.86$).

Koul, Clariana, and Salehi (2005) used the *ALA-Reader* sentence aggregate approach again to score students' essays on the structure and function of the heart and circulatory system. Working in pairs, participants researched this topic online and created concept maps using *Inspiration* software. Later, using their concept map, participants individually wrote a short essay. The concept maps and essays were scored by *ALA-Mapper* and *ALA-Reader* relative to an expert's essay, by another software tool called Latent Semantic Analysis (LSA), and by 11 pairs of human raters using two different rubrics. As in the previous study, the 11 human scores were averaged together into one composite essay score. Compared to the raters' composite score, the *ALA-Reader* scores were 5th out of 13, with an $r = .71$ (the 13 scores ranged from $r = .08$ to $.88$) and LSA scores were 9th out of 13, with an $r = .62$ indicating that *ALA-Reader* performed more like the average of the human raters than did LSA.

Clariana and Wallace (2007) used *ALA-Reader* to score essays relative to an expert referent and to establish and compare group average knowledge representations from those essays. As part of their final course examination, undergraduate business majors were asked to write a 300-word compare-and-contrast essay on four management theories that were covered during the course, a relevant and high stakes essay. The essays were scored by *ALA-Reader* using both a sentence and a linear aggregate approach. To serve as benchmarks, the essays were also separately scored by two human raters who obtained a Spearman rho inter-rater reliability of $\rho = .71$. The linear aggregate approach obtained larger correlations with the two human raters (ρ rater 1 = $.60$ and ρ rater 2 = $.45$) than did the sentence aggregate approach (ρ rater 1 = $.47$ and ρ rater 2 = $.29$). In addition, the group average network representations of low and high performing students were reasonable and straightforward to interpret, the high group was more like the expert, and the low and high groups were more similar to each other than to the expert.

In a follow up investigation, Clariana, Wallace, and Godshalk (2008) considered the effects of anaphoric referents on *ALA-Reader* text processing. Participants in an undergraduate business course ($N = 45$) again completed an essay as part of the course final examination. The investigators edited these essays to replace the most common pronouns "their", "it", and "they" with the appropriate referent. The original unedited and the edited essays for the top and bottom performing groups were processed with *ALA-Reader* using two different

approaches, sentence and linear aggregate. These data were then analyzed using Pathfinder analysis. The network average group representation similarity values comparing the original to the edited essays were large (i.e., about 90% overlap) but the linear aggregate approach obtained larger values than the sentence aggregate approach. The linear approach also provided a better measure of individual essay scores (e.g., $r = .74$ with composite rater scores).

These studies show a moderate correlation between human rater essay scores and *ALA-Reader* scores. Note that the essays in the first two studies used mostly technical biology vocabulary while essays in the third study used somewhat general vocabulary that included a number of synonyms for key terms, such as manager, supervisor, and boss for the key term 'management'. *ALA-Reader* may be more appropriate for some types of essays and is probably inappropriate for many types of essays. The more technical or specific the vocabulary in the essays, the better *ALA-Reader* should perform. In addition the first two studies used the sentence aggregate approach and obtained an adequate measure of essay performance, while in the third and fourth study, the linear aggregate approach provided a satisfactory measure of essay performance but the sentence aggregate approach did not. The linear approach appears to be better than the sentence wise approach, and this may relate to both the nature of structural knowledge as well as the forced linearity of expository text. As with *ALA-Mapper*, the evidence is persuading us that *ALA-Reader* is not an essay scoring tool, but rather it is a tool to measure knowledge structure and this measure of knowledge structure happens to correlate with various kinds of essay scores.

Next Steps

This chapter described two related software programs that were designed to complement Pathfinder analysis, *ALA-Mapper* for processing network graphs and *ALA-Reader* for processing text. The findings from several investigations were presented that indicate that these software tools may be measuring participants' knowledge structure. These two tools show potential but there are several critical issues yet to be resolved regarding these approaches.

Which key terms to use and how many should be used during *ALA-Mapper* and *ALA-Reader* analysis is an area for further investigation because some key terms appear to be far more important than others. Typically the course instructor or another content expert selects the key terms for the analysis. But further research must establish the best approach for determining the key terms. Contrary to expectations that using more terms means improved concurrent validity (Goldsmith et al., 1991), Clariana and Taricani (nd) used *ALA-Mapper* to score distance raw data from a set of 24 open-ended concept maps using either 16, 26 (16+10 more), or 36 (26+10 more) most important terms (as selected by a content expert). The largest correlations with the multiple choice terminology and comprehension posttests were observed for 16 terms, then 26, then 36. Increasing the number of terms used to score the concept maps did not increase the predictive ability of the scores, probably due to students not selecting enough of the most important words to include in their concept maps.

The kinds of information in open-ended concept maps (and network graphs) are very different than that in more closed concept maps. In open-ended concept mapping, students are typically only given a blank page and a prompt while closed mapping often also includes a list of terms, sometimes a list of linking phrases, and even in some cases a partially completed map with boxes for missing terms. This points out that concept mapping involves several different cognitive activities. The students' ability to generate the important terms is a critical task in open-ended concept mapping. Probably this generation task should be separated from the actual map formation

task by asking students to first list all terms that they would like to include in their map, and then in a second activity, provide a list of all important terms for the students to use. This two stage approach would maintain some of the strength of open-ended mapping (the gold standard approach) related to understanding the important concepts in a domain question while also requiring a full range of interaction with the actual important concepts, and also this two stage approach should produce more comparable maps thus reducing spurious measurement error.

Lambiotte et al. (1989, p. 342) proposed a taxonomy of ‘map devices’ based on the signalling device used to represent relationships among ideas: Spatially based, node based, link label based, and hybrid. The distance between terms in concept maps appears to be important information related to inference and comprehension (Cernusca, 2007, pp. 138-139; Clariana & Poindexter, 2003; Poindexter & Clariana, 2006; Taricani & Clariana, 2003, 2006), and so not only in network graphs but also the distances between key terms in a text passage may also be important information. A feature will be added to *ALA-Reader* to capture these linear distances between key terms in text as a proximity array in order to consider this notion.

Internal and external context factors strongly influence the kind of information elicited during concept mapping. For example, training participant’s to create hierarchical concept maps, whether the domain organization is hierarchical or not, must alter the obtained knowledge structure improperly towards hierarchical relationships. In a series of experiments, Derbentseva et al. (2007) showed that simply requiring participants to draw cyclic concept maps where clusters of four terms were connected in a circle with each leading to the next compared to tree (hierarchical) concept mapping resulted even in fundamentally different propositions. On average, 45% of the linking phrases between terms in the cyclic maps were dynamic phrases compared to only 14% of the linking phrases in the tree maps. Network graphs contain both associations (distances) and propositions (links), but a strong focus on either one by pre-training, the drawing prompt, or other context factors increases the information content of that aspect but at the expense of the other aspect. Many concept map investigations demand a strong emphasis on propositional correctness and the focus is so great that the distances between terms no longer have psychological meaning. In any case, strong context factors likely devastate the relationship between the artefact obtained and the participant’s actual knowledge structure.

Similarly, context variables that influence essays should be more closely examined to determine if context factors, such as the essay writing prompt, providing a list of terms, or the essay genre, can be manipulated to obtain essays that better capture the students’ structural knowledge. For example, compare-contrast type essay questions or other writing prompts which ask students to discuss the relationships between concepts may be most appropriate for eliciting knowledge structure (Goldsmith, Johnson, & Acton, 1991; Gonzalvo, Canas, & Bajo, 1994). Further research should consider what conditions best elicit essays that reflect students knowledge structure.

During the analysis phases, the referent data set and *PFNET* that is used as the baseline or standard to compare to the participants’ *PFNETs* is critical because error, idiosyncrasies, or spurious links in the referent *PFNET* produces error in every comparison. Referents must be carefully crafted, for example the linear and nonlinear referents described above. When expert’s concept maps or essays are used as the referent, probably several should be used since different experts may have different but correct representations of the domain question. At any rate, the approach used by investigators to obtain or create a referent must be carefully described, and if possible, the *PFNET* representation of that referent should be examined for under specification and for errors.

References

- Cernusca, D. (2007). *A design-based research approach to the implementation and examination of a cognitive flexibility hypertext*. Unpublished doctoral dissertation, University of Missouri, Columbia. Downloaded November 11, 2008 from <http://edt.missouri.edu/Summer2007/Dissertation/CernuscaD-071907-D8036/research.pdf>
- Clariana, R. B. (2002). *ALA-Mapper software*, version 1.01. Retrieved September 28, 2008 from <http://www.personal.psu.edu/rbc4/ala.htm>
- Clariana, R. B. (2004). *ALA-Reader software*, version 1.01. Retrieved December 24, 2008 from <http://www.personal.psu.edu/rbc4/score.htm>
- Clariana, R.B., & Koul, R. (2004). A computer-based approach for translating text into concept map-like representations. In A.J.Canas, J.D.Novak, and F.M.Gonzales, Eds., *Concept maps: theory, methodology, technology*, vol. 2, in the Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, Sep 14-17, pp.131-134. See <http://cmc.ihmc.us/papers/cmc2004-045.pdf>.
- Clariana, R.B., Koul, R., & Salehi, R. (2006). The criterion-related validity of a computer-based approach for scoring concept maps. *International Journal of Instructional Media*, 33, 317-325.
- Clariana, R.B., & Taricani, E. M. (n.d.). The consequences of increasing the number of terms used to score open-ended concept maps. *International Journal of Instructional Media*, 37 (2), in press.
- Clariana, R.B., & Wallace, P. E. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research*, 37, 209-225.
- Clariana, R.B., Wallace, P.E., & Godshalk, V.M. (2008). Deriving and measuring group knowledge structure via computer-based analysis of essay questions: the effects of controlling anaphoric reference. In Kinshuk, D. G. Sampson, J. M. Spector, P. Isaías, & D. Ifenthaler (Eds.), *Proceedings of the IADIS international conference on cognition and exploratory learning in the digital age* (88-95). Freiburg, Germany: International Association for Development of the Information Society.
- Dearholt, D.W. & Schvaneveldt, R.W. (1990). Properties of Pathfinder Networks. In Schvaneveldt (Ed.), *Pathfinder associative networks: studies in knowledge organization* (pp. 1-30). Norwood, NJ: Ablex Publishing Corporation.
- Derbentseva, N., Safayeni, F., & Canas, A.J. (2007). Concept maps: experiments on dynamic thinking. *Journal of Research in Science Teaching*, 44, 448-465.
- Einstein, G. O., McDaniel, M. A., Bowers, C. A. & Stevens, D. T. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 133-143.
- Goldsmith, T.E., & Davenport, D.M. (1990). Assessing structural similarity of graphs. In Schvaneveldt (Ed.), *Pathfinder associative networks: studies in knowledge organization* (pp. 75-87). Norwood, NJ: Ablex Publishing.
- Goldsmith, T. E., Johnson, P. J., & Acton, W.H. (1991). Assessing knowledge structure. *Journal of Educational Psychology*, 83, 88-96.
- Gonzalvo, P., Canas, J. J., & Bajo, M. (1994). Structural representations in knowledge acquisition. *Journal of Educational Psychology*, 86, 601-616.
- Harper, M. E., Hoefft, R. M., Evans, A. W. III, & Jentsch, F. G. (2004). Scoring concepts maps: Can a practical method of scoring concept maps be used to assess trainee's knowledge structures? *Factors and Ergonomics Society Annual Meeting Proceedings*, 48, 2599-2603.
- Jonassen, D.H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kinchin, I. M., & Hay, D. B. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42, 43-57.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychology Review*, 85, 363-394.
- KNOT (1998). *Knowledge Network and Orientation Tool for the Personal Computer*, version 4.3. Retrieved October 3, 2003 from <http://interlinkinc.net/>
- Koul, R., Clariana, R.B., & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. *Journal of Educational Computing Research*, 32, 261-273.
- Lomask, M., Baron, J. B., Greig, J., & Harrison, C. (1992, March). *ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science*. Paper presented at the annual meeting of the National Association of Research in Science Teaching, in Cambridge, MA.

- Poindexter, M. T., & Clariana, R.B. (2006). The influence of relational and proposition-specific processing on structural knowledge and traditional learning outcomes. *International Journal of Instructional Media*, 33, 177-184.
- Schvaneveldt, R. (1990). *Pathfinder associative networks: studies in knowledge organization*. Norwood, NJ: Ablex Publishing.
- Taricani, E. M. & Clariana, R.B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development*, 54, 61-78.