Chapter 7
Deriving Individual and Group Knowledge Structure from Network Diagrams and from Essays

Roy B. Clariana

7.1 Introduction

This chapter focuses on two somewhat fundamentally related ways to elicit knowledge structure, network diagrams (usually as concept maps), and essays (see Chapter 4, this volume). Because of the utility of the proven Pathfinder Network approach and its well-established research base, we developed software to convert concept maps and essays into data representations that can be analyzed with Pathfinder software. Our initial research focused on computer-based methods for scoring concept maps (e.g., Clariana, 2002) and since concepts 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 (Clariana & Koul, 2008; Koul, Clariana, & Salehi, 2005), and so 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 Pathfinder network analysis and then describes the *ALA-Mapper* and *ALA-Reader* scoring approach. Next the investigations with these two tools are reviewed, and finally suggestions for future research are provided.

7.2 Pathfinder Network Analysis

Pathfinder network analysis is a well-established system for deriving and representing the organization of knowledge (Jonassen, Beissner, & Yacci, 1993; Schvaneveldt, 1990). The Pathfinder algorithm converts estimates of relatedness of pairs of terms into a network representation of those terms called Pathfinder Networks (*PFNETs*) that are usually a two-dimensional representation of a matrix of relationship data in which concept terms are represented as nodes (also called

R.B. Clariana (✉)
The Pennsylvania State University, University Park, PA, USA
e-mail: rclariana@psu.edu

D. Ifenthaler et al. (eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge*, DOI 10.1007/978-1-4419-5662-0_7,

vertices) and relationships are shown as weighted links (also called edges) connecting the nodes. *PFNETs* resemble concept maps, but without link labels.

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 pairwise 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, the Knowledge Network and Orientation Tool for the Personal Computer software (KNOT, 1998) is used to reduce the raw proximity data into a *PFNET* representation that is a least-weighted path that links all of the terms. The set of links derived from Pathfinder analysis is determined from the patterns in the raw proximity data, and these are influenced by two parameters that can be manipulated by the researcher, q and Minkowski's r. These parameters for calculating the least-weighted path can be adjusted to reduce or prune the number of links in the resulting *PFNET* (refer to Dearholt & Schvaneveldt, 1990). The resulting *PFNET* 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 other 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 number of the links shared by two *PFNETs* (the intersection of two *PFNETs*). Similarity, which is also called neighborhood similarity, is the intersection divided by the union of two *PFNETs*.

Note that *KNOT* has a group average feature that can average 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 contained in individual *PFNETs*.

7.3 Network Diagrams and Knowledge Structure

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 (Taricani & Clariana, 2006).

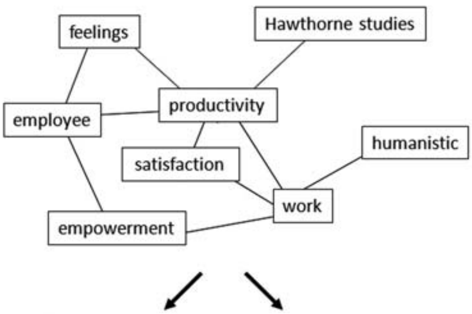
So what information in concept maps can be measured? There are at least three or four different cognitive tasks involved when creating a concept map that can leave a "cognitive residue" in the map. First, if the map is open-ended, where students may use any terms in their map, then a critical task is recalling (or possibly recognizing from a list) the most important terms/concepts to include in the 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 of associations. Then students identify propositions by linking pairs of terms with a line and adding a linking phrase to show 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 that likely reflects both the structure of their knowledge and an internalized graphic grammar or norm of what things like this should look like (Clariana & Taricani, 2010).

Probably the most fundamental meaningful psychological components of concept maps and essays are propositions (Einstein, McDaniel, Bowers, & Stevens, 1984; Kintsch, 1974; Kintsch & van Dijk, 1978). In essence, a proposition consists of a subject, a verb, and an object, which in a concept map consists of a term-(linking phrase)-term. *ALA-Mapper* converts node–node information in concept maps and other types of network diagrams into two separate kinds of raw data, links between terms and distances between terms measured in screen pixels (see Fig. 7.1).

Note that linking phrases may not be critical for concept map analysis. Harper, Hoeft, 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. Also, link labels are more computationally difficult to collect, handle, compare, and analyze than just the presence or absence of a link between terms. Thus *ALA-Mapper* pragmatically uses links only rather than matching link labels.

7.3.1 ALA-Mapper Investigations

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 completing 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 distances between terms in the concept maps and to represent the links that connected terms (see Fig. 7.1). Using Pathfinder *KNOT* software, the raw distance and link data were converted into *PFNETs* and



	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work
empowerment	–							
employee	1	–						
feelings	0	1	–					
Hawthorne studies	0	0	0	–				
humanistic	0	0	0	0	–			
productivity	0	1	1	1	0	–		
satisfaction	0	0	0	0	0	1	–	
work	1	0	0	0	1	1	1	–

Link array

	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work
empowerment	–							
employee	66	–						
feelings	97	50	–					
Hawthorne studies	166	160	120	–				
humanistic	143	161	135	57	–			
productivity	91	87	60	77	77	–		
satisfaction	52	68	66	114	98	39	–	
work	114	143	127	83	34	68	76	–

Distance array

Fig. 7.1 A sample network diagram and its link array and distance array

then were compared to an expert's *PFNET* to obtain network 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) *ALA-Mapper* link-based scores were 0.36, (b) *ALA-Mapper* 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) *ALA-Mapper* link-based scores were 0.76 and (b) *ALA-Mapper* 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 applied to posttest network diagrams (e.g., no linking phrases) rather than concept maps. The mapping directions specifically directed the participants to use spatial closeness to show relationships and intentionally deemphasized the use of links. Participants completed one of three print-based text lesson treatments 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 then completed three multiple-choice posttests that assessed identification, terminology, and comprehension and were finally asked to draw a network diagram given a list of 25 pre-selected terms. The adjunct question lesson treatment was significantly more effective than the other lesson treatments for the comprehension outcome, and no other treatment comparisons were significant. *ALA-Mapper* network diagram scores based on link data were more related to terminology ($r = 0.77$) than to comprehension ($r = 0.53$), while *ALA-Mapper* network diagram scores based on distance data were slightly more related to comprehension ($r = 0.71$) than to terminology ($r = 0.69$). It was suggested that the links drawn to connect terms related to verbatim knowledge from the lesson text covering facts, terminology, and definitions; while the distances between terms in the network diagram related to comprehension of the processes and functions of the heart and circulatory system.

In a follow-up study, Taricani and Clariana (2006) asked 60 undergraduate students to read a print-based instructional text on the heart and circulatory system and then create concept maps of that content. Half of the participants were given feedback in the form of a prepared hierarchical concept map and the other half did not receive this 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 ($r = 0.78$) than to comprehension ($r = 0.54$) whereas concept map scores derived from distance data were more related to comprehension ($r = 0.61$) than to terminology ($r = 0.48$).

This supports the idea that there is worthwhile relational information in the distances between terms in a network diagram. 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 map damages the distance information captured in the map. For example, concept map training that demands that all map elements be propositions that are term-(linking phrase)-term direct 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, the Poindexter and Clariana (2004) investigation described above provided mapping directions that intentionally deemphasized 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.

7.3.2 Rubrics and Network Diagram Scores

So far, the investigations above have avoided the issue of what precisely are raters scoring when they score a concept map or other type of network diagram. Those studies above used holistic scoring that only considered the content accuracy reflected in the concept maps, a quantitative approach where more correct ideas obtains a higher score form the raters. However, Koul, Clariana, and Salehi (2005) reported that *ALA-Mapper* data correlated better with raters' scores using a qualitative than a quantitative rubric. In their investigation, teachers enrolled in a graduate course worked in pairs to research a science topic online and then 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 *ALA-Mapper* and *ALA-Reader* and by human raters using qualitative and quantitative rubrics. The quantitative rubric was adapted from the Lomask, Baron, Greig, and Harrison (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 structures which may be interpreted as indicators of progressive levels of understanding: (1) *Spoke*, a structure in which all of 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 nearby; and (3) *Net*, a network both highly integrated and hierarchical, demonstrating a deep understanding of the topic. *ALA-Mapper* concept map scores were a 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. We hold that this measure of knowledge structure is tapping a fundamental level of knowledge, the association network that can be drawn from to create meaningful propositions on the fly. Also, distance data and link data in network diagrams can both contain interesting and useful information. However, internal and external context factors can enhance or suppress the information content in this raw data and so must be well controlled. So our focus turned to measuring knowledge structure and on refining the writing prompts to elicit better knowledge representations.

7.4 Essays and Knowledge Structure

The *ALA-Reader* essay analysis approach was adopted directly from the *ALA-Mapper* network diagram analysis approach. *ALA-Reader* searches for key terms in text that are then represented as links in an aggregate array either (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 diagram link raw data (see the bottom left portion of Fig. 7.1) 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) although any text genre is likely influenced by the writer'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.

7.4.1 Sentence Aggregate Approach

The *ALA-Reader sentence aggregate approach* was developed to analyze text at the sentence level because sentences are an important unit of text organization. Sentences contain one or more propositions and the sentence aggregate 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 pre-selected key terms (and their synonyms and metonyms). 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 $(n^2 - n)/2$ elements. Each cell in the array corresponds to a pair of key terms (see the left panel of Fig. 7.2). A "1" entered in the appropriate cell of the array indicates that two key terms co-occurred in the same sentence and a "0" indicates that those two key terms did not occur in the same sentence. The software continues to aggregate sentences into the array until all of the text is processed. For example, given the

<u>Sentence 1</u>

	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work
empowerment	–							
employee	0	–						
feelings	0	0	–					
Hawthorne studies	0	0	0	–				
humanistic	0	0	0	0	–			
productivity	0	0	0	0	1	–		
satisfaction	0	0	0	0	1	1	–	
work	0	0	0	0	1	1	1	–

<u>Sentence 1, 2, 3, & 4</u>

	empowerment	employee	feelings	Hawthorne studies	humanistic	productivity	satisfaction	work
empowerment	–							
employee	1	–						
feelings	0	1	–					
Hawthorne studies	0	1	0	–				
humanistic	0	0	0	0	–			
productivity	1	1	1	1	1	–		
satisfaction	0	0	0	0	1	1	–	
work	1	1	0	0	1	1	1	–

Fig. 7.2 The sentence aggregate proximity file created by *ALA-Reader* for sentence 1 (*left panel*) and for all four sentences (*right panel*)

following 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 translated by *ALA-Reader* into this sentence aggregate array of selected key terms shown in Fig. 7.2. The force directed network diagram of the four sentences is shown in Fig. 7.3.

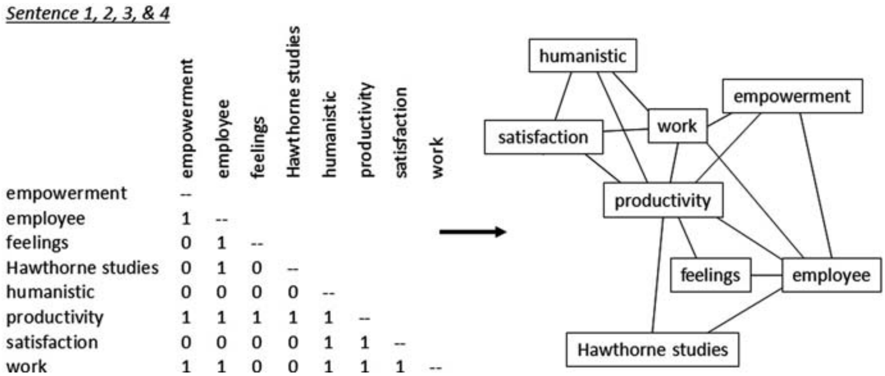


Fig. 7.3 The force-directed graph of the four-sentence aggregate

7.4.2 Linear Aggregate Approach

In contrast to the sentence aggregate approach, the *ALA-Reader linear aggregation approach* enters a "1" (1s) in the appropriate cell of the array to represent adjacent key terms during a linear pass through the text, and so will always obtain a connected graph. However, 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 diagram begins to fold bringing related terms closer together in the two-dimensional space. The same text used above in Fig. 7.2 would be reduced by *ALA-Reader* to this linear sequence of selected key terms (with link numerical order shown here for clarity): "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 Fig. 7.4 and its *PFNET*). The linear sequence of terms can also be represented as a force-directed graph (a *PFNET*)

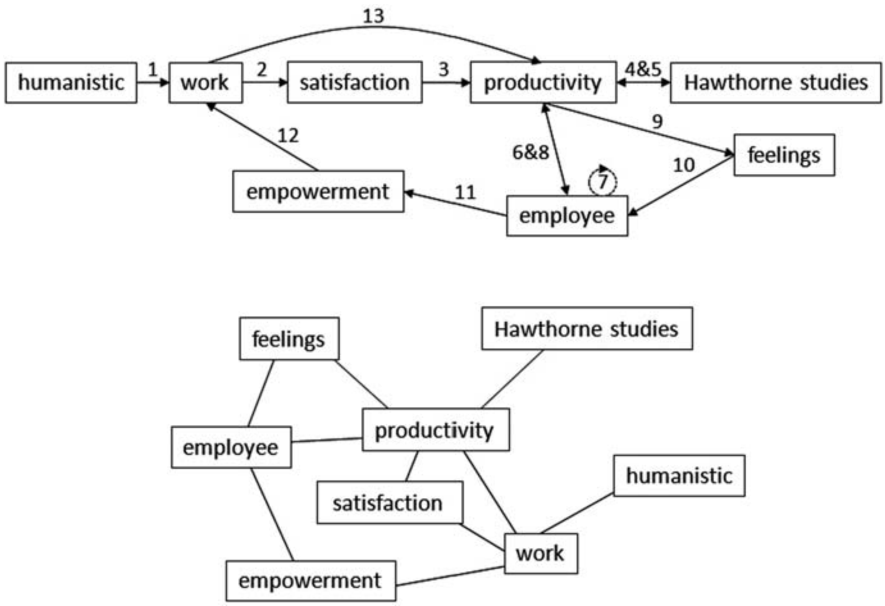


Fig. 7.4 A visual representation of the linear sequence of key terms from a participant's essay passage about the Humanistic management approach (*top panel*) and its equivalent force-directed graph (*PFNET, bottom panel*)

that highlights 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 the bottom panel of Fig. 7.4). For example, the key term "productivity" with five links is a high-degree node (i.e., with three 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. Also, compare the number and pattern of links for the sentence aggregate *PFNET* in Fig. 7.3 to the linear aggregate *PFNET* of the same text shown in Fig. 7.4 to note the difference in *ALA-Reader* representation output between the sentence and linear approaches.

Clariana and Koul (2004) used *ALA-Reader* software to score 12 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 combined together into one composite essay score, then all scores were correlated with the human raters' composite score. Compared to the composite score, the *ALA-Reader* scores were 5th out of 12, with an $r = 0.69$ (the 12 scores correlations ranged from $r = 0.11$ to 0.86), which indicates that four raters were better than the *ALA-Reader* sentence aggregate scores, but eight raters were worse. Also, the *ALA-Reader* scores were not included when creating the composite score, and so this is a conservative comparison that strongly favors the raters' scores.

Koul et al. (2005) used the *ALA-Reader* sentence aggregate approach 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 by *ALA-Reader* relative to an expert's map and 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 rater scores were averaged together into one composite essay score (a conservative value that favors the raters). Compared to the composite essay score, the *ALA-Reader* essay scores were 5th out of 13, with an $r = 0.71$ (the 13 scores ranged from $r = 0.08$ to 0.88) and *LSA* scores were 9th out of 13, with an $r = 0.62$. As before, relative to the rater composite score, *ALA-Reader* performed better than eight of the raters and also was better than *LSA* on this specific biology content essay.

Clariana and Wallace (2007) used *ALA-Reader* to score essays on management theories relative to an expert referent and also to establish and compare group average knowledge representations derived from those essays. As part of their final course examination, undergraduate business majors ($N = 29$) were asked to write a 300-word compare-and-contrast essay on four management theories from 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 $\rho = 0.71$. The linear aggregate approach obtained larger correlations with the two human raters (ρ rater $1 = 0.60$ and ρ rater $2 = 0.45$) than did the sentence aggregate approach (ρ rater $1 = 0.47$ and ρ rater $2 = 0.29$). 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 both 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, with a Pearson correlation $r = 0.74$ with the raters' composite score.

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 second two used fairly 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 may be inappropriate for many types of essays. In these few studies, the more technical or specific the vocabulary in the essays, the better *ALA-Reader* performed. In addition, the first two studies used the sentence aggregate approach only 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 approach, and this may relate to both the nature of structural knowledge and the forced linearity of expository text. As with *ALA-Mapper*, the evidence is persuading us that *ALA-Reader* is not really 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.

7.5 Next Steps

This chapter described two related software programs that were designed to complement Pathfinder analysis, *ALA-Mapper* for processing 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.

A critical area for further investigation is which key terms to use and how many should be used during *ALA-Mapper* and *ALA-Reader* analysis 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 phase. But further research must establish the best approach for determining these key terms. Contrary to expectations that using more terms means improved concurrent validity (see Goldsmith et al., 1991), Clariana and Taricani (2010) used *ALA-Mapper* to score distance data from a set of 24 open-ended concept maps using either 16, 26 (those 16 + 10 more), or 36 (those 26 + 10 more) most important terms (as selected and prioritized by a content expert). The greatest 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.

Another area for further research involves the effects of providing participants with the key terms during concept mapping or when writing their essay. In open-ended concept mapping, students are typically 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 blank boxes for missing terms. The different approaches involve different cognitive activities (e.g., levels of generativity; Lim, Lee, & Grabowski, 2008). A students' ability to recall 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 researcher-selected terms for the students to use during actual mapping. This two stage approach would maintain some of the power of open-ended mapping (the gold standard) related to understanding the important concepts in a domain question, while also requiring a full range of interaction with the concepts during the second stage.

Another critical area for further research is the setup and prompt used for eliciting a concept map or an essay. Internal and external context factors strongly influence the kind of information elicited during concept mapping. For example, training participants to create hierarchical concept maps, whether the domain organization is hierarchical or not, must alter the obtained knowledge structure improperly toward hierarchical relationships. In a series of experiments, Derbentseva, Safayeni, and Canas (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 diagrams 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 artifact 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 et al., 1991; Gonzalvo et al., 1994). Further research should consider what conditions best elicit essays that reflect student's knowledge structure.

The referent used for comparison analysis also requires considerable thought and further research. 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* produce error in every comparison. Referents should be carefully crafted. 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 then be inspected for under specification and for errors.

Also, there are more than two approaches (i.e., linear and sentence) for translating essays into arrays. Lambiotte et al. (1989, p. 342) proposed a taxonomy of “map devices” based on the signaling 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, 2006), and so not only in network diagrams 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.

In summary, the history of science has shown that new observation tools lead to different ways to conceptualize phenomenon, and this leads to new and more powerful theories. The software tools described in this chapter and in this volume show considerable promise for the systematic analysis of knowledge.

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, & 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. (2008). The effects of learner prior knowledge when creating concept maps from a text passage. *International Journal of Instructional Media, 35*, 229–236

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., & Poindexter, M. T. (2003). *The influence of relational and proposition-specific processing on structural knowledge.* A paper presented at the Annual Meeting of American Educational Research Association (AERA), San Diego, CA, April 2003.

Clariana, R. B., & Taricani, E. M. (2010). The consequences of increasing the number of terms used to score open-ended concept maps. *International Journal of Instructional Media, 37*(2), 218–226.

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 D. G. Kinshuk, J. M. Sampson, P. Spector, D. Isaías, & D. Ifenthaler (Eds.), *Proceedings of the IADIS international conference on cognition and exploratory learning in the digital age* (pp. 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., Hoeft, 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*. Hillside, 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.

Lambiotte, J. G., Dansereau, D. F., Cross, D. R., & Reynolds, S. B. (1989). Multirelational semantic maps. *Educational Psychology Review, 1,* 331–367.

Lim, K. Y., Lee, H. W., & Grabowski, B. (2008). Does concept-mapping strategy work for everyone? The levels of generativity and learners' self-regulated learning skills. *British Journal of Educational Technology, 40*, 606–618.

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.