Structure and Substance in Digital Communication

David A. Thomas
Mathematics Education Associates, LLP
david.cynthia.thomas@gmail.com

Livia Sa
University of Nevada, Reno
liviasa@unr.edu

Zhongxiao Li
Georgia Gwinnett College
zli@ggc.usg.edu

Cleborne D. Maddux
University of Nevada, Reno
Maddux@unr.edu

Abstract

A growing trend in post-secondary education is the supplementing of classroom discourse with online discourse. This new discourse is often situated in social networking and course management applications. There are several reasons for the move toward online discourse. Wide availability of new hardware and software has played a role. Even more important is an emerging consensus that our societal, economic, and individual futures depend on our capacity to engage complex issues in a collaborative manner. However, this new technology and the techniques they make possible have introduced new problems. Many instructors are finding it more difficult to monitor, evaluate, and direct class discourse. *Social network analysis* and *content analysis* can provide the means to understand the structure of online, digital discourse and the content of digital documents.

Increasingly, post-secondary educators are supplementing classroom discourse with online discourse situated in social networking applications and course management tools such as *WebCT*, *Blackboard,* and *Moodle*. Professionals in many disciplines are using these applications to facilitate and promote student collaboration while encouraging knowledge discovery and improvement. Driving this change is an emerging consensus that our societal, economic, and individual futures depend on our capacity to engage complex issues, challenges, and opportunities in a collaborative manner.

To achieve this in the case of STEM education, we must do more than challenge students to master the content of their chosen disciplines. We must help them develop the communication and collaboration skills that will be increasingly important to modern professional and R&D subcultures.

Paradoxically, the more students adopt these empowering practices and technologies, the more difficult it is for instructors to monitor, evaluate, and direct class discourse. For instance, monitoring an online discussion that generates 20 messages per week may require nothing more than a few hours reading and reflection on the part of the instructor. But what happens when an online discussion generates hundreds of messages per week? Under these circumstances, even conscientious instructors are likely to miss the emergence of student leaders (or isolates) and powerful new ideas and/or misconceptions. New strategies and supporting technologies are needed that empower instructors to meet this challenge.

Part I of this paper reviews the literature relative to collaborative knowledge development and improvement in online communities. Part II presents a conceptual overview of social

To be published in the Journal of Educational Technology Systems, Volume 38, Number 3.

network analysis (SNA) and discusses the use of *NetMiner*, a powerful SNA technology, to study the structure of online discourse such as that found in threaded discussions. Part III presents a conceptual overview of content analysis (CA) and discusses the use of *Leximancer*, a versatile text mining technology, to study the substance of digital documents. Part IV presents examples of the use of both social network analysis and content analysis to understand the structure and substance of digital communication in web-assisted courses.

Part I: Collaborative Knowledge Creation and Improvement in Online Communities

Community of Inquiry Model

Garrison (2006) emphasizes the importance of reflection and discourse in online learning, scientific inquiry, and other collaborative pursuits. Common forms of online asynchronous communication include email, blogs, and threaded discussion forums. In face-to-face environments, verbal agility, spontaneity, and confidence may determine who speaks and who listens. Online discourse, on the other hand, may be motivated and facilitated in ways that promote thoughtful reflection and systematic discourse. Indeed, online asynchronous communication may be more effective than face-to-face discourse at promoting collaboration, discourse, and community among university students. For instance, asynchronous communication may result in more "group-centered" interaction patterns while face-to-face communication is frequently more "authority-centered" (Lobel, Neubauer, & Swedburg, 2005).

A learning theory that is situated in asynchronous learning environments is the Community of Inquiry model (Arbaugh, 2007; Garrison, Anderson, & Archer, 2000). In this model, teaching and learning occur in an environment characterized by cognitive presence, social presence, and teaching presence (see Figure 1).

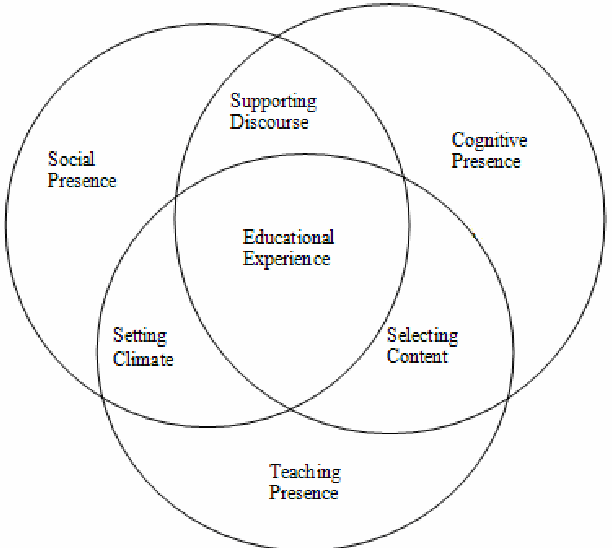

Figure 1: Community of Inquiry Model

In the Community of Inquiry model, cognitive presence represents the extent to which the participants are able to construct meaning through sustained communication. Social presence represents the ability of learners to project their personal characteristics into the community of inquiry, thereby presenting themselves as 'real people.' Teaching presence represents the design,

facilitation, and direction of cognitive and social processes for the purpose of realizing learning outcomes.

This model is consistent with a social constructivist view of learning (Lapadat, 2002) since it posits that knowledge is not handed down by instructors, but is constructed by students as they engage course content and one another in discourse. In asynchronous learning networks (ALNs), informal conversations and other social behaviors can be used to create and maintain a sense of community (Hoadley & Pea, 2002). These kinds of conversations contribute to the social presence of the ALN. However, if the main purpose of an ALN is to enhance cognitive presence, mechanisms are needed to ensure that the social aspects of discourse do not detract from or supplant the cognitive agenda. Research on ALNs has shown that there are problems related to the quantity and quality of online interactions. These problems can undermine inquiry goals (van Aalst, 2006).

With regard to quantity, Guzdial and Turns (2000) studied 35 online university courses and reported an average of only 4.8 notes (messages) sent per student over 10 weeks. With regard to quality, Hewitt (2003) found that students rarely revisited a note once they had read it and they rarely returned to a discussion thread once they had read most of the notes in that thread. Similarly, online discussions have been found to focus primarily on facts rather than causal explanations (Hakkarainen, Lipponen & Järvelä, 2002; Lipponen, 2000). Because of these and other findings, it may be helpful to rethink the purpose and nature of the work that students do online, moving from a focus on discussion to an emphasis on a variety of collaborative activities (van Aalst, 2006). Then too, this may imply that we should consider shifting the emphasis of social psychology research to the question of how to enable groups to achieve greater quality in problem solving (Witte, 2007).

Knowledge Creation and Improvement

A key concept in the development of new models of online learning is that of distributed cognition (Karasavvidis, 2002). Rogers (2005) describes distributed cognition as a theory involving the interactions between people, artifacts, and internal and external representations. Pea (1993) argues that the concept of intelligence should not be thought of as a property of mind alone and should include the artifacts and tools that we use to represent complex thoughts and to alleviate tedious and burdensome cognitive tasks. In this context, collaborative knowledge creation may be viewed as a process occurring in the distributed cognition of a team of human collaborators and all of the representations, tools, and other technical aides used. The effect of this frame of reference is to emphasize the importance of the manner in which ideas are shared, refined, extended, and tested. As an example, Pea (1993) argued that with the help of imaging software, computers can be used to augment the skill of visualization and consequently contribute to the understanding of complex relationships in ways that would be impossible without them. In order to facilitate collaborative knowledge creation, strategies and tools are needed that make clear at any time what has been learned, what is of current interest, and what opportunities for inquiry remain. Strategies might include note development procedures that summarize previous ideas and introduce new concepts, reuse of notes for new purposes, and personalizing or annotating the emerging knowledge base.

In other words, a new conceptualization of online work is needed. In this new conceptualization, discussion should not play the central role. Instead, emphasis would be on a variety of activities intended to iteratively improve the quality and utility of ideas developed in a learning community (van Aalst, 2006).

For instance, in scientific inquiry a fundamental goal is to advance the frontier of knowledge. This requires the creation and improvement of ideas new to the discipline. Scardamalia and Bereiter (2003) call this type of inquiry 'knowledge building.' There is growing evidence that students in elementary school and above are capable of participating in knowledge building in ways that develop their natural curiosity (Bereiter, 2002; Scardamalia, 2002). An important question is whether students perceive the discourse in which they are engaged as a method for producing new knowledge. In other words, do they treat ideas as objects of inquiry that can be improved by scrutiny, debate, testing and modification (Bereiter, 2002)? If they do, students may also recognize that disciplinary knowledge, such as the content of textbooks has been produced through a similar process. Rather than seeing expert knowledge as certain and complete, they may learn to see it as improvable. Such positions are consistent with the highest stage of King and Kitchener's Reflective Judgment Model, a model that describes the development of critical thinking in adolescence and adulthood (King & Kitchener, 1994).

Collaborative Teams

Collaboration is critical because team cognitive properties and processes amount to more than the sum of the properties and processes of the individuals that comprise that team (Hutchins, 1991; Hinsz, Tindale, & Vollrath, 1997; Levine, Resnick, & Higgins, 1993). Researchers have suggested that the team itself may be considered a cognitive or information processing unit (Hinsz et al., 1997), another expression of the concept of distributed cognition in which team coordination and communication are critical aspects of team performance (Brannick & Prince, 1997). Witte (2007) identifies several strategies shown to improve team performance:

- The higher the quality of individual input regarding the subject matter at the beginning of the group interaction, the higher the quality of the group performance (Grofman, 1978; Lorge & Solomon, 1955; Sorkin, Hays, & West, 2001);
- The more individual inputs are independent of one another at the beginning of the group interaction, the higher the quality of the group performance (Sorkin et al., 2001).
- The more the group increases individual independent inputs, the higher the quality of the group performance (Hinsz et al., 1997);
- The more comprehensible the individual input is for each group member, the higher the quality of the group performance (Libby, Trotman, & Zimmer, 1987);
- The more high quality individual input influences the final group decision, the higher the quality of the group performance (Littlepage, Schmidt, Whisler, & Frost, 1995).

Witte goes on to note that normative influences on group performance (e.g., "group think;" "majority wins") processes can manipulate and distort the integration of individual input (Hinsz et al., 1997) and should therefore be reduced to a minimum. Consequently, collaborative teams should strive to:

- maximize the informational influence on the group performance process;
- minimize the normative influence on the group performance process; and
- optimize the influence of individual input on the final group decision.

In this approach, the role of the facilitator is to ask leading questions such as "What additional data would you need to explore that conjecture?" or "How could you justify that conclusion?" Facilitators may also provide information on general principles and practices of interest to all of the teams. They should not "help" teams as they struggle with their respective research hypotheses, designs, analyses, and document drafts. This suggestion is echoed by Simonsen and Banfield (2006).

In summary, the improvement of online discourse in collaborative settings should take into account the following research findings:

- students given appropriate training are capable of collaborative inquiry;
- collaborative learning should motivate and facilitate team discourse focused on knowledge development and improvement;
- knowledge development and improvement in collaborative teams is best served when individual participants contribute their insights and expertise continuously;
- technology-based collaboration tools are needed that facilitate the reuse of information by team members as knowledge emerges and evolves in team discourse;
- knowledge construction may be thought of as occurring in a distributed knowledge, information, and computing space encompassing the individual team members and all of the information resources, modeling tools, and artifacts employed.

Part II: Social Network Analysis and the Structure of Online Discourse

In asynchronous learning networks (ALNs), information flows between students and teachers in a variety of formats including email, chatrooms, threaded discussions and documents. Researchers seeking to understand the dynamics of this exchange and its associated cognitive and social outcomes often focus solely on either the structure or the content of the communication. For instance, in the process of communicating with one another in asynchronous threaded discussions, it has been found that students create implicit networks of relations with one another (Aviv, Erlich, Ravid, & Geva, 2003). Typically, communication networks include emergent structural features that appear and evolve over time with different individuals playing different roles. These roles have been studied and described for years by researchers who have specialized in social network analysis. For example, in responding to one another, it has been found that some students play a role known as *information broker*. Those who play the role of *coordinator* facilitate within their own study group. Those who are *gatekeepers* receive communication from other groups and pass it along to their own group members. *Representatives* pass on ideas to other groups, while *itinerants* act as facilitators between two or more individuals of another group. *Liaisons* facilitate communication between individuals in two groups.

By studying the connectivity structures of these networks, researchers using highly time-intensive manual methods have probed the underlying mechanisms that establish and maintain the networks, making possible the development of theories that explain their emergence (Contractor, N. S., Wasserman, S., & Faust, K., 2006; Monge, P. R. and Contractor, N. S., 2003). Social network analysis (SNA) has provided both a visual and a mathematical context for analyzing the structure of asynchronous online communication (Wasserman & Faust, 1994). New hardware and software tools have great potential for accelerating the process of understanding social networks in online discourse. In fact, social network analysis and these new tools are now being used to analyze online collaborative learning communities (Aviv, Erlich, & Ravid, 2005) and to correlate network structures with knowledge construction (Aviv, Erlich, Ravid, & Geva, 2003).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	15.0	7.0	10.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	20.0	0.0	5.0	6.0	17.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	6.0	3.0	0.0	2.0	7.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	8.0	4.0	1.0	0.0	19.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	6.0	13.0	4.0	18.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	1.0	0.0	0.0	0.0	2.0	0.0	8.0	5.0	11.0	7.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	1.0	0.0	7.0	0.0	4.0	4.0	6.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	1.0	0.0	7.0	3.0	0.0	5.0	5.0	0.0	0.0	0.0	0.0	0.0
9	1.0	0.0	1.0	0.0	0.0	4.0	3.0	5.0	0.0	12.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	1.0	0.0	9.0	8.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	2.0	1.0	2.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	3.0	3.0	3.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.0	0.0	2.0	2.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	2.0	0.0	2.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.0	1.0	2.0	0.0

Table 1: Fictional Response Data

Network relations may be represented using a student response relation matrix, illustrated using the fictional data seen in Table 1. In this table, cell *(i,j)* indicates how many times the *i*-th (row) learner responded to postings by the *j*-th (column) learner during asynchronous communications. For example, learner 3 responded 6 times to the postings of learner 1. Notice that this matrix is not symmetrical. That is, not all learners engage one another in the same ways. Data of this sort may be obtained directly from server logs and/or compiled reports generated by discussion forum software.

Graphical representations of this sort of data may be obtained using SNA software, such as *NetMiner* (Cryam, 2007). Figure 2 shows how *NetMiner* visualizes the data presented in Table 1. In this visualization, gender is coded by shape (~~i.e.~~ circles for males and triangles for females), assigned discussion groups are coded by color (~~i.e.~~, group 1 in red and group 2 in blue), and age is coded by size (older students appear as larger shapes). This visualization shows that three female students (blue triangles # 8, 9, and 10) participated in the male discussion group (red) rather than their assigned female group. It also shows that some students communicated with 6 or more of their peers (e.g., #1, 3, 4, 6, 7, 9), while every student communicated with at least 4 peers. [Insert Figure 2 about here]

In addition to response data, student attribute data may also be incorporated in analyses. Table 2 shows the gender, group assignment, examination score, project score, and age of the same 15 fictional students seen in Table 1. This data may be analyzed using well-established and traditional statistical methods. In the case of many social network analysis tools, these analyses may be conducted within the same data management and analysis environment used to create the visualization seen in Figure 2. For instance, *NetMiner* (Cryam, 2007) may be used to test the null hypothesis that examination performance is unrelated to gender. The output of that analysis is seen in Table 3 and the result provides no evidence for the rejection of a null hypothesis ($p > .05$). These and other data visualization and analysis features make *NetMiner* a powerful tool for investigating the structure of online discourse.

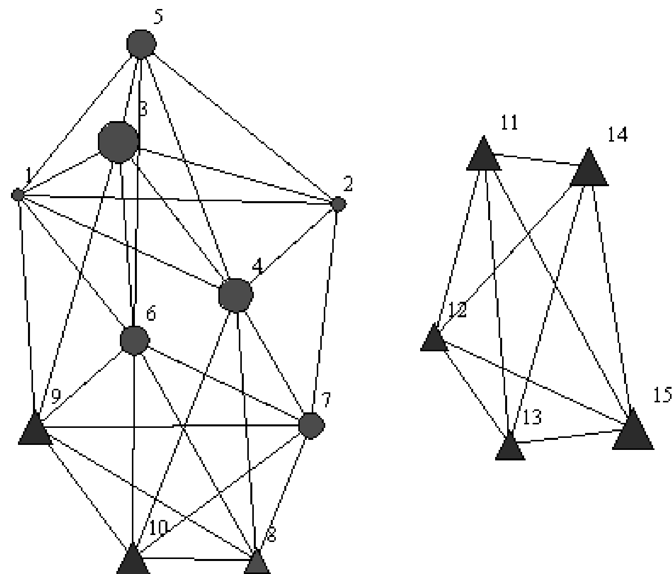

Figure 2: Structure of Student Discourse

	Gender	Groups	Exam	Project	Age
1	1.0	1.0	75.0	1.0	20.0
2	1.0	1.0	80.0	1.0	24.0
3	1.0	1.0	92.0	1.0	61.0
4	1.0	1.0	50.0	1.0	50.0
5	1.0	1.0	78.0	1.0	44.0
6	1.0	1.0	80.0	1.0	42.0
7	1.0	1.0	87.0	1.0	37.0
8	2.0	1.0	79.0	2.0	39.0
9	2.0	2.0	70.0	1.0	50.0
10	2.0	2.0	55.0	1.0	52.0
11	2.0	2.0	99.0	1.0	49.0
12	2.0	2.0	100.0	1.0	40.0
13	2.0	2.0	87.0	1.0	44.0
14	2.0	2.0	80.0	2.0	54.0
15	2.0	2.0	68.0	1.0	60.0

Table 2: Fictional Attribute Data

The authors of this paper have prepared a Website that is a demonstration of a few of the extremely basic capabilities of NetMiner. This site can be found at http://www.scs.unr.edu/~maddux/socdis. The site can be accessed by using the password "social." The site contains sample analyses and explanations as well as links to some hypothetical data derived from student email.

Statistics: ANOVA - Vector

ANALYSIS DATA OVERVIEW

TITLE	C:\Documents and Settings\dthomas3\Desktop\Cleb Data\sampledata1.xls		
ATTRIBUTE VARIABLES	Categorical Vector : Gender Continuous Vector : Exam		
# OF NODES	15		
TRANSFORM HISTORY			
C > \Documents and Settings\dthomas3\Desktop\Cleb Data\sampledata1.xls			

RESULT OF ANOVA VECTOR

Source of variation	Sum of Squares	d.f.	Mean Square
Between groups	20.119	1	20.119
Within groups	2,755.214	13	211.94
Totals	2,775.333	14	

Normal Test

Eta-Square	F ratio	p
0.007	0.095	0.763

Table 3: *NetMiner* Analysis of Attribute Data

Part III: Content Analysis and the Substance of Digital Discourse and Documents

While it is important to understand *how* information flows between students and teachers, it is even more important that instructors understand *what* is being communicated. C*ontent analysis* (also called *discourse analysis* in some contexts) provides methodologies for studying the substance of communication. Thanks to technological advancements, today's scholars have access to vast quantities of textual information such as articles, books, web pages, and blogs. For the most part, these data are unstructured rather than structured (an example of unstructured data is that found in journal articles in contrast to structured data, an example of which is that found in databases)**.** Structured data is generally analyzed using data mining and statistical methods familiar to many scholars. Unstructured data, on the other hand, must be treated using text mining and content analysis methods unfamiliar to many scholars. Figure 3 contrasts these approaches to data retrieval and analysis.

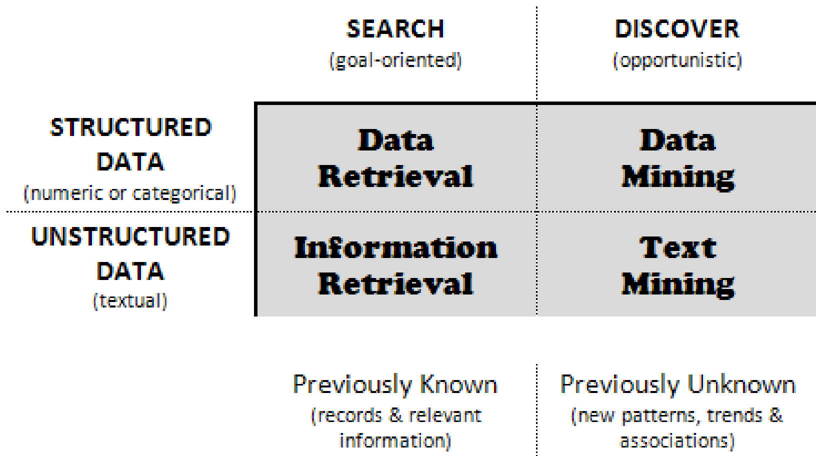

Figure 3: Structured vs. Unstructured Data

The exponential growth of unstructured data resources and the need to analyze these resources on a practical time scale in an unbiased manner has motivated the development of a variety of text-mining methodologies and technologies. Text-mining today is a computer assisted methodology that extracts knowledge from large amounts of information (Feldman & Sanger, 2007; Weiss, Indurkhya, Zhang, & Damerau, 2005). For example, Sa (2009) used *Leximancer*, a computerized text mining tool, to analyze the content of a body of mathematics education literature consisting of 1,202 abstracts from 7 professional journals (See Figure 4).

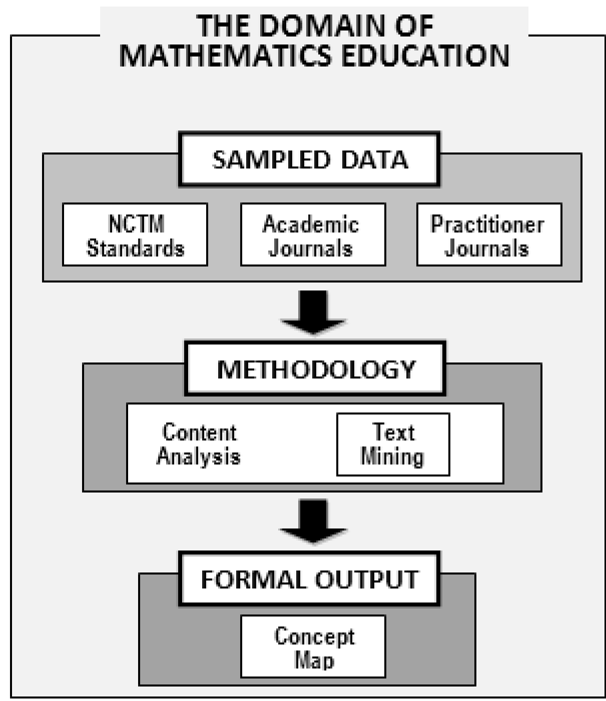

Figure 4: Content Analysis of Mathematics Education Literature

The steps in this process are illustrated in Figure 5. After uploading the raw data,

Leximancer transformed the text into a carefully structured, intermediate data format. Knowledge discovery algorithms were then applied to this representation of the data (Feldman & Sanger, 2007), resulting in a table of word frequencies and a matrix of their relative co-occurrence. Next, frequently occurring words were identified as seeds and co-occurring seeds were aggregated into concepts. Finally, a clustering algorithm was used to produce a concept map, which was then edited and interpreted by the researcher. In this approach, the resulting concept map is a result of both automated and expert analysis.

An advantage of automated processes such as this is that they provide perfect coder reliability in the application of coding rules to text. According to Weber (1990, p. 41), algorithmically derived analysis offers "high coder reliability … freeing the investigator to concentrate on other aspects of inquiry, such as validity, interpretation, and explanation."

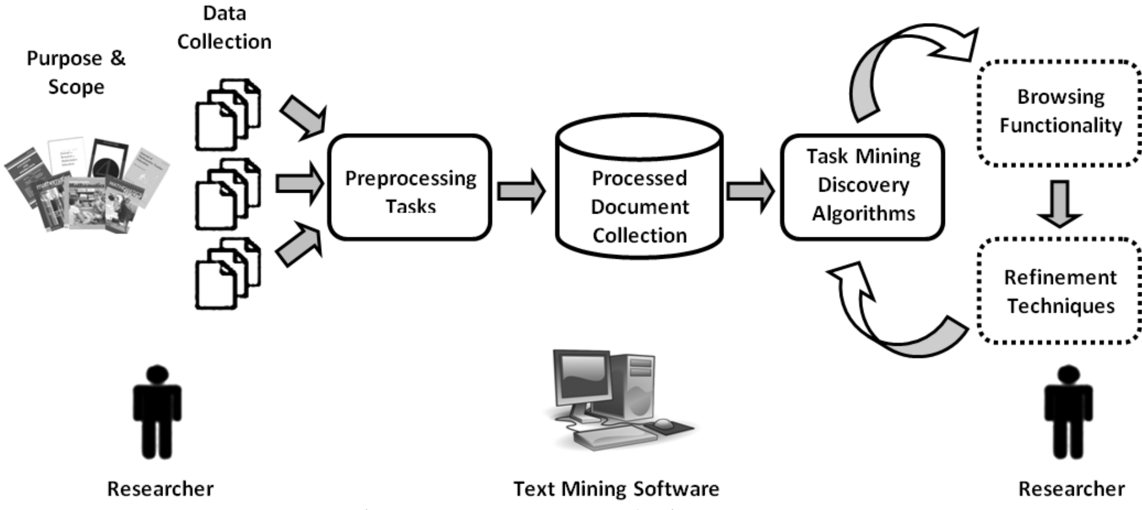

Figure 5: Content Analysis Process

The resulting concept map provides three sources of information about the content of the document set: (a) the main concepts contained within the text and their relative importance; (b) the strengths of links between concepts (how often they co-occur); and (c) the similarities in the context in which they occur. These findings may be used to compare documents (e.g., to find documents similar to a given document), to characterize communication patterns or practices (e.g., to determine if there are gender-based differences in the content of student communication), and to construct an overview of a large body of information. In this concept map:

- the brightness of a concept is related to its frequency (i.e. the brighter the concept, the more often it appears in the text);
- the brightness of links relate to how often the two connected concepts co-occur closely within the text;
- nearness in the map indicates that two concepts appear in similar conceptual contexts (i.e. they co-occur with similar other concepts).

Figure 6 identifies the three main themes spanning the document set: (a) students; (b) research, and (c) education. Figure 7 overlays the concepts associated with these themes and tags associated with the publications represented in the document set. Note the relative proximity of the concept *research* and the tag representing the *Journal for Research in Mathematics*

Education (TG_3_JRME_TG) and the relative separation from the tag representing the practitioner-oriented journal *Mathematics Teaching in the Middle School* (TG_6_MTMS_TG). Figures 8 and 9 provide insights into the co-occurrence of the concept *research* with many other concepts in the document set. *Leximancer* provides direct links via the book icons in Figure 9 to these co-occurrences in the original text (See Figure 10). These links provide context for detailed analysis of these occurrences by the researcher.

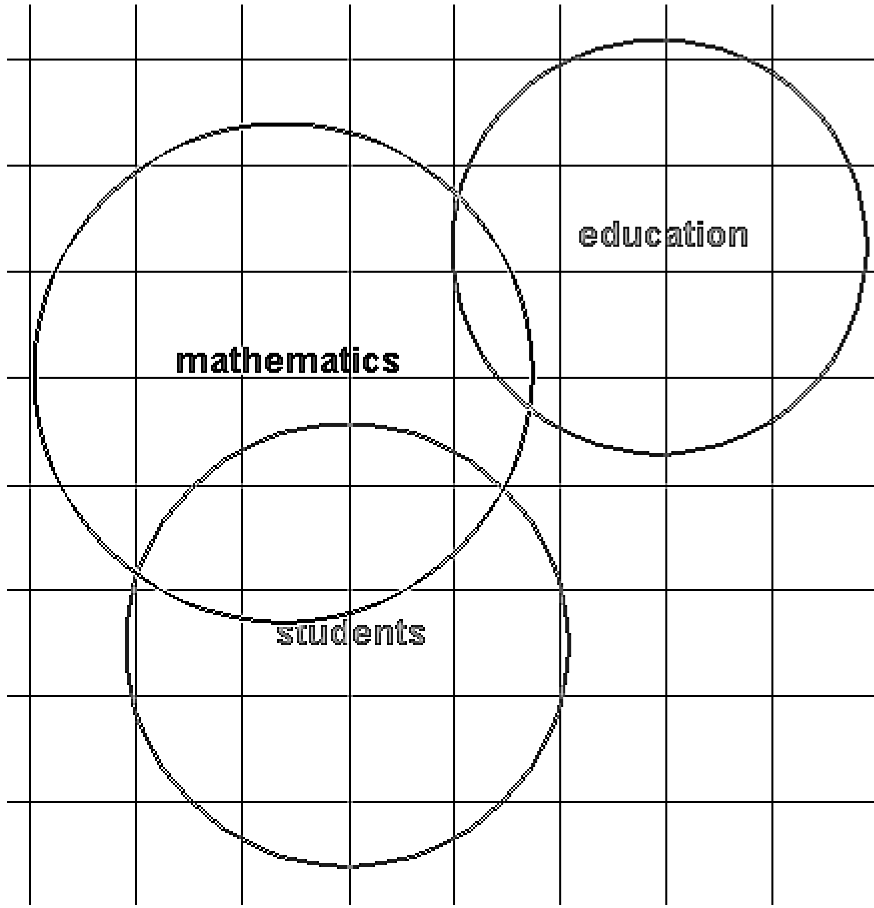

Figure 6: Main Themes

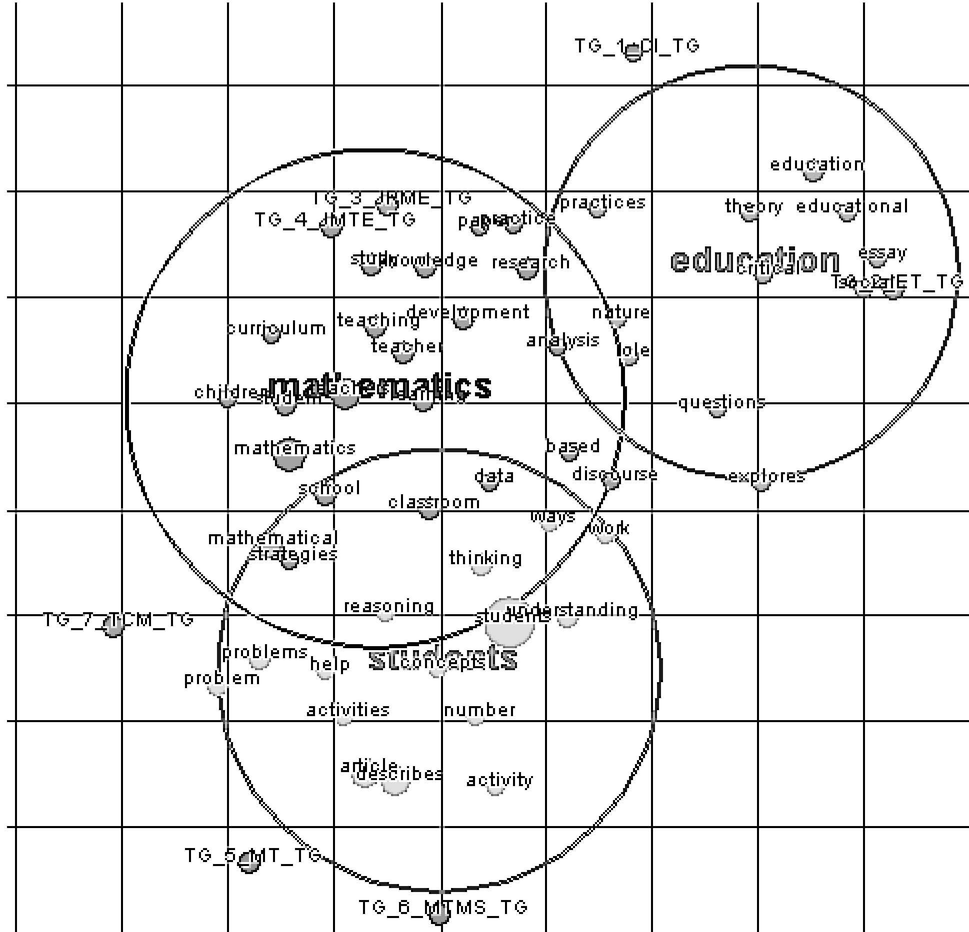

Figure 7: Concepts within Themes and Journal Tags

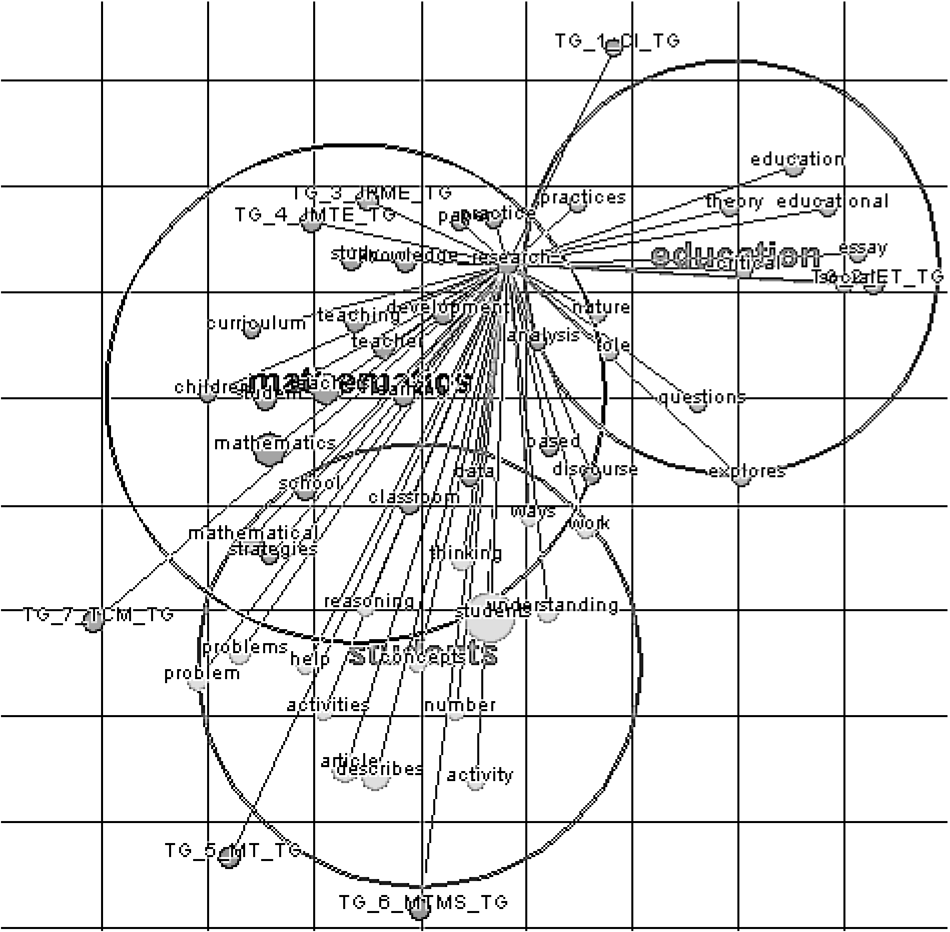

Figure 8: Map of Concepts Co-occurring with *research*

Related Entities (and locations). (Count: 54)

Click on the **buttons** to browse the evidence

Concept	Absolute Count	Relative Count	
students	110	77.4%	
mathematics	75	52.8%	
teachers	56	39.4%	
describes	39	27.4%	
TG_3_JRME_TG	38	26.7%	
article	36	25.3%	
education	35	24.6%	

Figure 9: Partial List of Concepts Co-occurring with *research*

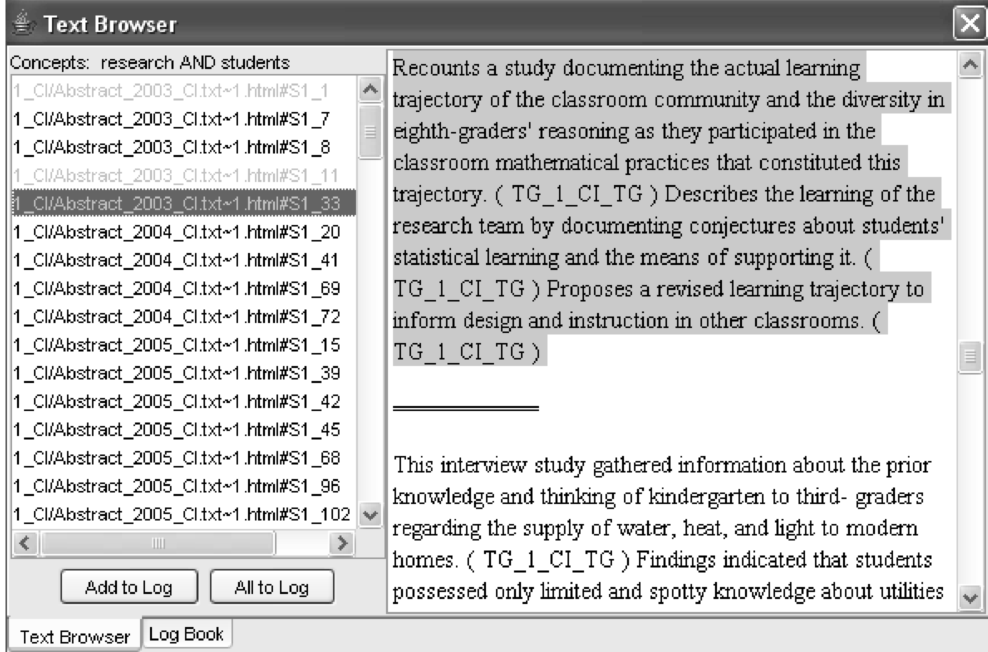

Figure 10: Links to Use of the concept *research* in the Original Text

These and other *Leximancer* features may be used to identify main ideas and clusters of ideas within large document sets and to point the researcher to all occurrences of a concept or phrase. Applied to online student discourse, this capability may be used to identify critical concepts and misconceptions and to associate these with individual students or teams of students.

Part IV: Examples of SNA and CA of Student Discourse and Documents

Thomas, Li, Knott, & Li (2008) used SNA to analyze the structure of over 1500 messages spanning three undergraduate, web-assisted mathematics content courses for preservice elementary and secondary teachers. In these courses, students were assigned to small collaborative groups the first day of class. Every 3 weeks, each group submitted a lengthy homework report to the instructor. The content of these homework reports and the content of the course examinations were closely aligned. Homework reports were developed in each group's *WebCT* discussion forum and graded periodically by the instructor. Homework items with errors and/or omissions received no credit. In other words, group member had to read, edit, and revise each others' work or suffer the consequences. This practice also motivated students to work closely with teammates rather than with members of other groups. Accepting and assuming this responsibility was not easy for many students. In time, the burden of the consequences overcame students' reluctance and every team developed an authentic collaboration. The authors call this sort of collaboration *forced*.

In each course, some students participated frequently, while others were relatively silent. Some students played facilitating roles, while others contributed little. These and other structural aspects of student discourse were strongly related to individual student achievement on course examinations (for example, an r^2 statistic of 0.55 was calculated between measures of student participation in the network and measures of their achievement on course examinations). In other words, without examining what students actually said to each other, the structure of their discourse strongly suggested which students were succeeding and which were struggling. Figure 11 illustrates the structure of student communication in the three courses. In this figure, male students are represented with small icons and female students with large icons.

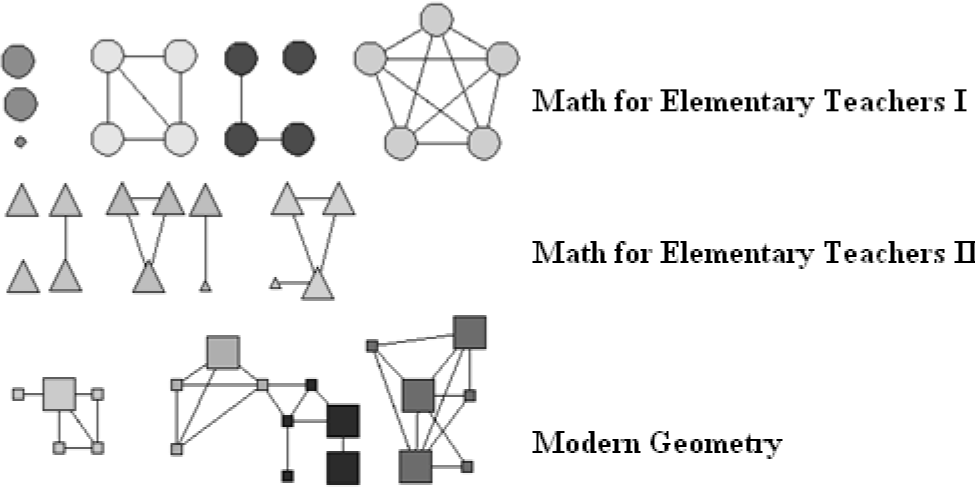

Figure 11: Structure of Student Communication in Three Mathematics Courses

In a follow-up study involving another section of Mathematics for Elementary Teachers I, Li (2009) removed the forcing mechanisms used in the first study. In this approach, which the authors call *unforced*, online student discourse focused on topical discussions rather than graded homework reports. These discussions had a much smaller relationship to their grades than the homework reports in the prior course. Finally, while students were seated as collaborative groups during class, they were free to read and respond to all postings in the online discussion forum.

Figure 12 strongly suggests that the structure of student communication was independent of group membership (dotted rectangles). A statistical analysis confirmed that the structure of student discourse under these conditions was not related to student achievement on course examinations, even though a cursory examination of the data suggest a tendency to that effect. These two studies suggest that while the structure of online discourse may be associated with student achievement in some circumstances, the strength of that relationship depends strongly on the purpose and structure of the learning tasks and related motivational factors.

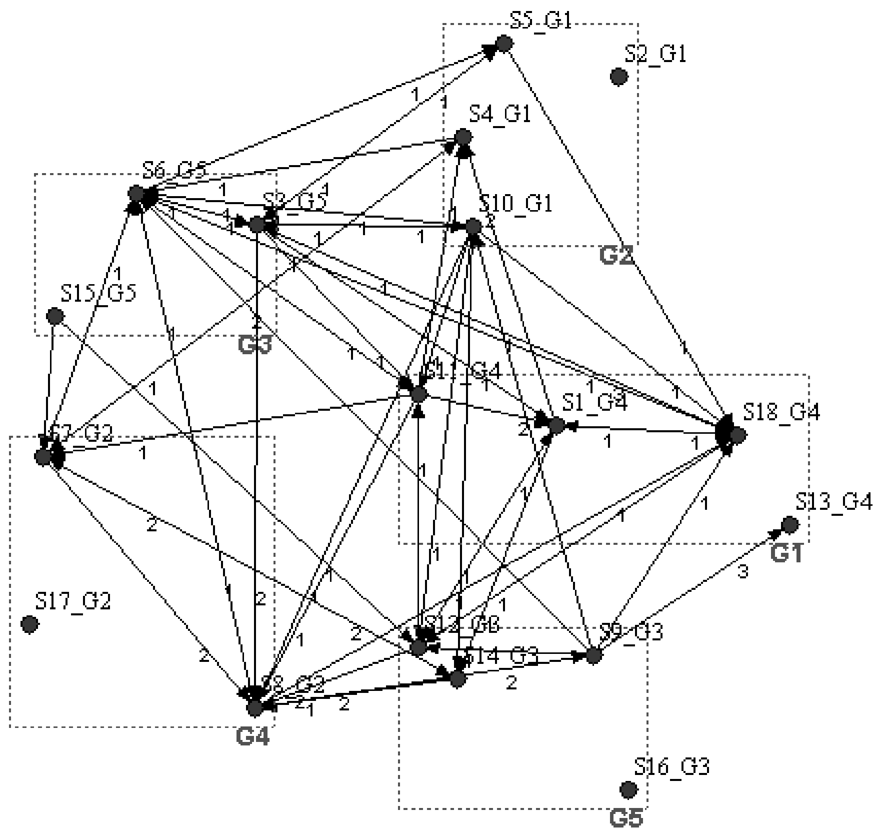

Figure 12: Communication Between Participants

Li's study (2009) also investigated the content of student communication. His methodology included both a traditional hand-scored, rubric-based approach to data analysis and a machine-scored approach using *Leximancer* software. The rubric-based approach relied on the judgment of a team of mathematics educators to rate the mathematical merit of individual postings. *Leximancer* was used to create an overall perspective of student discourse. The research questions addressed and principal findings in this analysis follow.

- How consistent is the level of discourse for individual students?

Figures 13, 14, and 15 show the percentage of high, medium, and low level messages posted by students during the first third, middle third, and final third of the course. Clearly, students did not post consistently at any level.

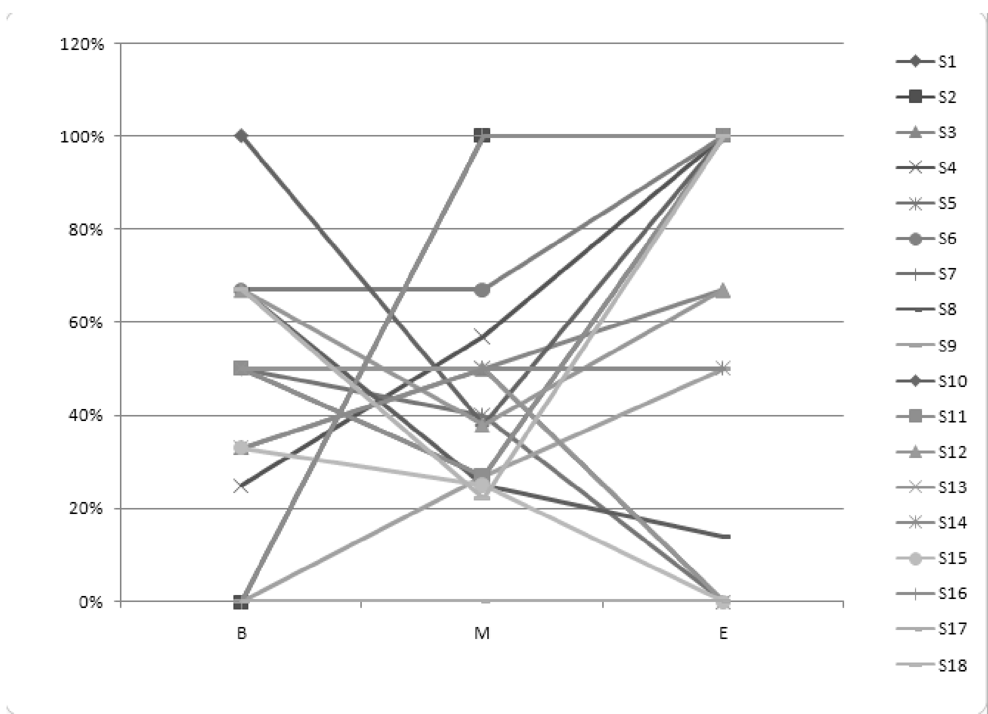

Figure 13: Percentage of High Level Messages by Individual

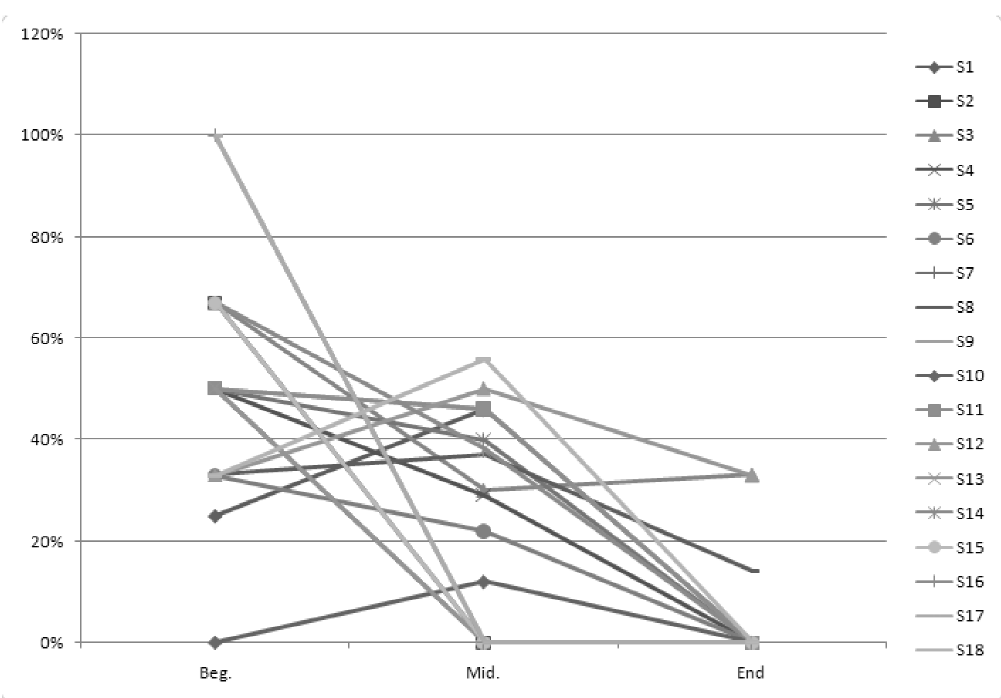

Figure 14: Percentage of Middle Level Messages by Individual

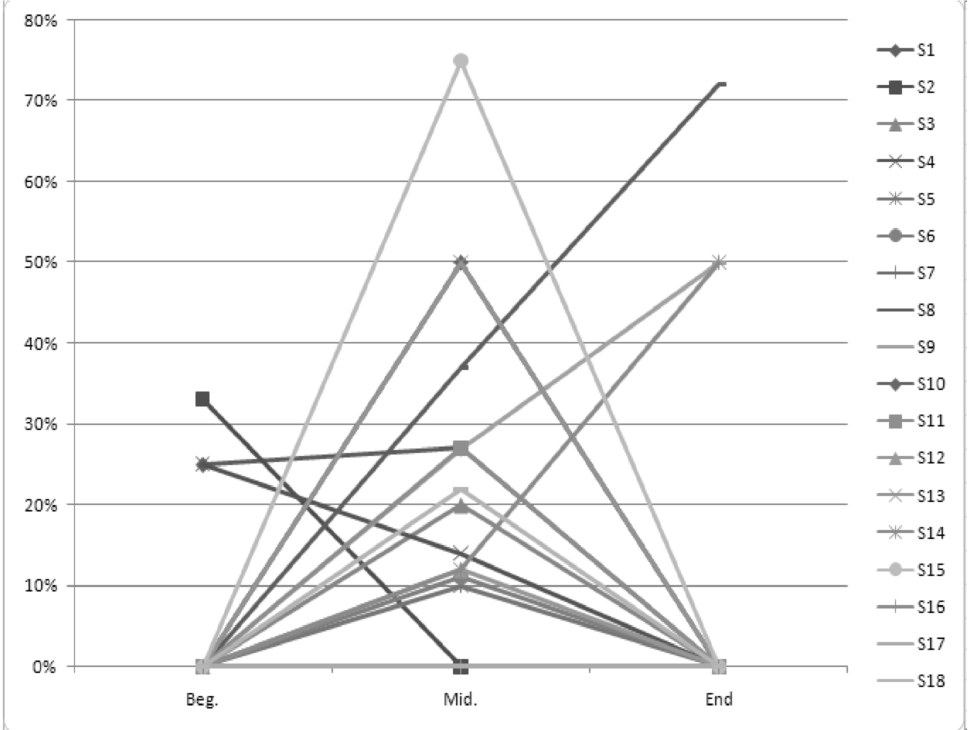

Figure 15: Percentage of Low Level Messages by Individual

- How consistent is the level of discourse over time from topic to topic?

As shown in Figure 16, the level of student discourse from topic to topic increased over time as fewer low and medium level messages were posted relative to high level messages.

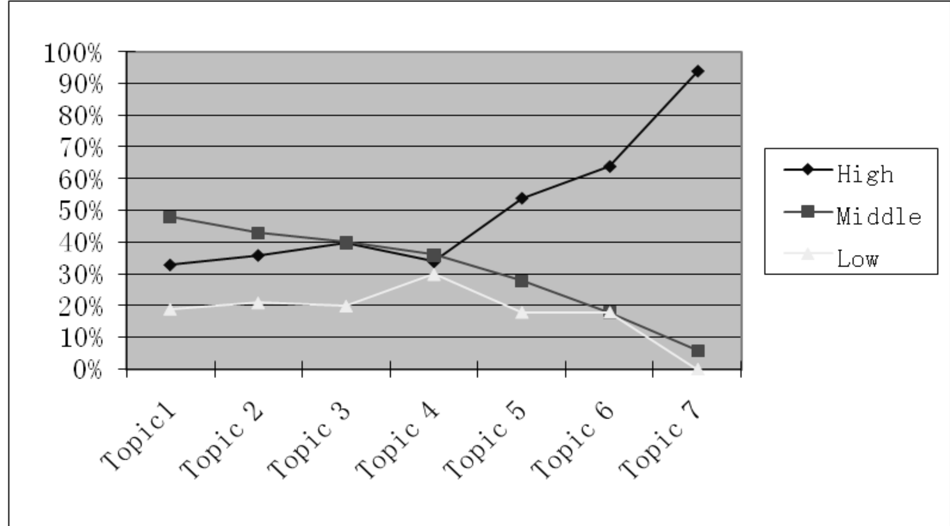

Figure 16: Change of the Percentage of High/Medium/Low Level Messages Over Topics

- How are role centrality and concept centrality related in student discourse?

Using social network analysis, Li (2009) identified the central player(s) in each topical discussion (for example, the student(s) with the highest *role centrality*). Using content analysis,

he identified the key concepts in each discussion such as *concept centrality*. As shown in Table 4, the student(s) with the highest role centrality were the same students who contributed the ideas with the highest *content centrality*.

Topics	Central Player(s)	Contributors of Central Concepts
1 &2	S1	S1
3 &5	S1 & S18	S1 & S18
4	S14	S14
6 & 7	S9	S9

Table 4: Central Players in the Network vs. Central Concept Contributors

A similar but slightly weaker relationship was found relative to students whose role centrality was secondary to that of the students identified in Table 5.

Topics	Secondary Player(s)	Contributors of Central Concepts
1 &2	**S3, S4, S8**, S14,	**S3**, **S4, S8**, S6, S16
3 &5	**S3, S8,** S9, S12,	**S3,** S6, **S8**
4	**S10**	S3, S1, S6, **S10,** S18
6 & 7	**S8, S14, S16**	S1, **S8, S14, S16**

Table 5: Secondary Players in the Network vs. Central Concept Contributors

These findings suggest a strong relationship between the leadership roles that students play in a network and the quality of their contributions to student discourse. A necessary caution is that this analysis does not deal with the causality of this relationship, though further content analysis may shed light on that issue.

Discussion

The capacity of instructors to improve online discourse is dependent on their understanding of its evolving structure and content. *Social network analysis* and *content analysis* provide methodological and technological means for identifying and characterizing evolving structural and substantive aspects of student communication. Highly refined techniques have been worked out by researchers over the years using time-intensive manual methodologies. Recent advancements in hardware and software have automated many of these techniques and made them accessible to researchers in other domains. Timely information of this sort might be used by instructors to reassign struggling or discontented individuals to different online groups, to reconfigure groups and/or tasks to optimize emerging capabilities, and to form a leadership team consisting of individuals from each student group. Use of social network analysis and content analysis to study online communication in web-based courses is in its infancy. Nevertheless, new research methodologies and technologies crafted to meet the challenges of online learning are beginning to emerge that will empower both researchers and students.

Among the findings discussed in this paper, the following seem instructive when designing and delivering web-assisted and web-based courses in which student discourse is used to promote knowledge discovery and improvement.

- If students are to develop authentic, productive collaborations, strong incentives are needed;

- If students are to develop and improve knowledge, challenging, structured tasks are needed;
- When strong incentives and challenging, structured tasks are provided, strong associations are possible between the structure of student discourse and student achievement;
- Role centrality (e.g., who is leading) and content centrality (e.g., what is being said) are related. That is, student leaders are often the strongest contributors of high-quality content.

There is another, less obvious advantage to the mixed approach advocated in this paper. Over the years, the philosophical/methodological gulf between proponents of quantitative and qualitative research methodologies has widened to the point that meaningful dialogue between these "camps" is rare. For instance, "qualitative" researchers interested in teaching and learning in online environments have relied heavily on rubric-based, hand-coding procedures and qualitative interpretive practices to analyze online discourse. While this approach has produced a large body of work, much of it is unsatisfying to "quantitative" researchers. Likewise, many "qualitative" researchers find quantitative studies unsatisfying. We believe that technology-based *social network analysis* and *content analysis,* both of which have solid statistical foundations, may provide "common ground" on which quantitative and qualitative researchers may, once again wrestle with important questions and ideas without surrendering the strengths of their research methodologies.

Assuming that a suitable forum emerges for exploring mixed methods of content and structural analysis of online, asynchronous, student discourse, opportunities for collaborative research may arise naturally among the growing, nationwide community of faculty interested in improving teaching and learning in web-assisted and web-based courses. If that happens, it seems likely that many research questions normally associated with traditional instruction may be extended to web-based and web-assisted instruction. For instance,

- To what extent do web-based learning environments advantage/disadvantage students whose learning styles, language skills, and personalities interact with web-based instructional modalities?
- What skills and dispositions do teachers and students need to succeed in web-based, collaborative learning environments?
- How do instructors use content and structural analysis of online student discourse to assess student progress and to inform instructional decisions?
- In what sort of courses might this sort of analysis be most useful? Least useful? Why?

From the perspective of the authors, these investigations will do more than extend the scope of educational research; they add a new and distinctive "flavor" to what we do as scholars.

References

Arbaugh, J. (2007). An empirical verification of the Community of Inquiry framework. *Journal of Asynchronous Learning Networks*, 11(1).

Aviv, R., Erlich, Z. and Ravid, G. (2005). Response neighborhoods in online learning networks: A quantitative analysis. *Educational Technology & Society,* 8(4).

Aviv, R., Erlich, Z., Ravid, G. and Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks,* 7(3).

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bloome, D., Carter, S.P., Christian, B.M., Otto, S. & Shuart-Faris, N. (2005). *Discourse analysis & the study of classroom language & literacy events - A microethnographic perspective.*

Mahwah, NJ: Lawrence Erlbaum Associates.
Bluesine Consulting (2005). *Leximancer manual*. Cambridge, MA: Bluesine Consulting.
Brannick, M., & Prince, C. (1997). An overview of team performance measurement. In M.T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theories, methods, and applications*. Mahwah, NJ: Erlbaum.
Chimombo, M., & Roseberry, R.L. (1998). *The power of discourse: An introduction to discourse analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
Contractor, N., Wasserman, S., & Faust, K. (2007). *Testing multi-level, multi-theoretical hypotheses about networks in 21 century organizational forms: An analytic framework and empirical example*, http://www.spcomm.uiuc.edu/users/nosh/manuscripts/pstarpaper.html.
Cryam (2007). *NetMiner II, Ver. 2.6.0,* Seoul: Cryam Co., Ltd. http://www.netminer.com.
Garrison, D. (2006). Online collaboration principles. *Journal of Asynchronous Learning Networks,* 10(1).
Garrison, D., Anderson, T., & Archer, W. (2000). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education* 2(2).
Grofman B. (1978). Judgmental competence of individuals and groups in a dichotomous choice situation: Is a majority of heads better than one? *Journal of Mathematical Sociology*, *6*.
Guzdial, M. & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, *9*.
Hakkarainen, K., Lipponen, L. & Järvelä, S. (2002). Epistemology of inquiry and computer supported collaborative learning. In T. Koschmann, R. Hall & N. Miyake (Eds), *CSCL 2: Carrying forward the conversation*. Mahwah, NJ: Lawrence Erlbaum Associates.
Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, *28*(1).
Hiltz, S. R. & Goldman, R. (2005). *Learning together online: research on asynchronous learning networks*. Mahwah, NJ: Lawrence Erlbaum.
Hinsz, V., Tindale, R., & Vollrath, D., (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, 121, 43 -64.
Hoadley, C. & Pea, R. D. (2002). Finding the ties that bind: tools in support of a knowledge building community. In K. A. Renniger & S. Wesley (Eds), *Building virtual communities: learning and change in cyberspace*. Cambridge, UK: Cambridge University Press.
Hutchins, E. (1991). The social organization of distributed cognition. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, DC: American Psychological Association.
Karasavvidis, I. (2002). Distributed cognition and educational practice. *Journal of Interactive Learning Research*, 13(1/2).
King, P. & Kitchener, K., (1994). *Developing reflective judgment: understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass Publishers.
Lapadat, J.(2002). Written interaction: A key component in online learning. *Journal of computer-mediated communication, 7(4).*
Levine, J., Resnick, L., & Higgins, E., (1993). Social foundations of cognition. *Annual Review of Psychology, 44,* 585 -612.
Leximancer (2008). www.leximancer.com.
Li, Z. (2009). *Asynchronous Discourse in a Web-assisted Mathematics Education Course.*

Unpublished doctoral dissertation, University of Idaho.
Libby, R., Trotman, K., & Zimmer, I. (1987). Member variation, recognition of expertise, and group performance. *Journal of Applied Psychology*, 72(1).
Lipponen, L. (2000). Towards knowledge building: from facts to explanations in primary students' computer mediated discourse. *Learning Environments Research*, *3*, 179–199.
Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input process–output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, *69*, 877–889.
Liu, L. & Maddux, C. (in press). Web 2.0 articles: Content analysis and a statistical model to predict recognition of need for new instructional design strategies. *Computers in the Schools*.
Lobel, M., Neubauer, M., & Swedburg, R., (2005). Selected topics from a matched study between a face-to-face section and a real-time online section of a university course. *International Review of Research in Open and Distance Learning*, 6(2).
Lorge J., & Solomon, H. (1955). Two models of group behavior in the solution of Heureka-type problems. *Psychometrica*, *20*, 139–148.
Monge, P. & and Contractor, N. (2003). *Theories of communication networks*, Oxford, UK: Oxford University Press, 2003.
Pea, R., (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). New York: Cambridge University Press.
Redmond, P. & Lock, J., (2006). A flexible framework for online collaborative learning. *The Internet and Higher Education*, 9, 267-276.
Rogers, Y. (2005). Distributed Cognition and Communication. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics* (2nd edn). Oxford: Elsevier.
Sa, L. (2009). *Content analysis of unstructured data in the field of mathematics education using text mining*. Unpublished doctoral dissertation, University of Nevada, Reno.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago, IL: Open Court.
Scardamalia, M. & Bereiter, C. (2003). Knowledge building. In J. W. Guthrie (Ed.), *Encyclopedia of education, second edition* (pp. 1370–1373). New York, NY: Macmillan Reference.
Simonsen. L. & Banfield, J. (2006). Fostering mathematical discourse in online asynchronous discussions: An analysis of instructor interventions. *Journal of Computers in Mathematics and Science Teaching*, 25(1).
Sorkin, R. D., Hays, C., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review*, *108*, 183–203.
Thomas, D., Li, Q., Knott, L., & Li, Z. (2008). The structure of student dialogue in web-assisted mathematics courses. *Journal of Educational Technology Systems*, 36(4), 415-431.
van Aalst, J. (2006). Rethinking the nature of online work in asynchronous learning networks. *British Journal of Educational Technology*, 37(2).
Wasserman, S. & Faust, K. (1994). Social Network Analysis. Methods and applications. Cambridge: Cambridge University Press.
Weber, R.P. (1990). *Basic content analysis*. Newbury Park, CA: Sage Publications.
Witte, E. (2007). Toward a group facilitation technique for project teams. *Group Processes & Intergroup Relations*, 10(3).