Epistemic Network Analysis and Topic Modeling for Chat Data from Collaborative Learning Environment

Zhiqiang Cai
The University of Memphis
365 Innovation Drive, Suite 410
Memphis, TN, USA
zcai@memphis.edu

Brendan Eagan
University of Wisconsin-Madison
1025 West Johnson Street
Madison, WI, USA
eaganb@gmail.com

Nia M. Dowell
The University of Memphis
365 Innovation Drive, Suite 410
Memphis, TN, USA
niadowell@gmail.com

James W. Pennebaker
University of Texas-Austin
116 Inner Campus Dr Stop G6000
Austin, TX, USA
pennebaker@utexas.edu

David W. Shaffer
University of Wisconsin-Madison
1025 West Johnson Street
Madison, WI, USA
dws@education.wisc.edu

Arthur C. Graesser
The University of Memphis
365 Innovation Drive, Suite 403
Memphis, TN, USA
art.graesser@gmail.com

ABSTRACT

This study investigates a possible way to analyze chat data from collaborative learning environments using epistemic network analysis and topic modeling. A 300-topic general topic model built from TASA (Touchstone Applied Science Associates) corpus was used in this study. 300 topic scores for each of the 15,670 utterances in our chat data were computed. Seven relevant topics were selected based on the total document scores. While the aggregated topic scores had some power in predicting students' learning, using epistemic network analysis enables assessing the data from a different angle. The results showed that the topic score based epistemic networks between low gain students and high gain students were significantly different ($t= 2.00$). Overall, the results suggest these two analytical approaches provide complementary information and afford new insights into the processes related to successful collaborative interactions.

Keywords

chat; collaborative learning; topic modeling; epistemic network analysis

1. INTRODUCTION

Collaborative learning is a special form of learning and interaction that affords opportunities for groups of students to combine cognitive resources and synchronously or asynchronously participate in tasks to accomplish shared learning goals [15; 20]. Collaborative learning groups can range from a pair of learners (called a dyad), to small groups (3-5 learners), to classroom learning (25-35 learners), and more recently large-scale online learning environments with hundreds or even thousands of students [5; 22]. The collaborative process provides learners with a more efficient learning experience and improves learners' collaborative learning skills, which are critical competencies for students [14]. Members in a team are different in many ways. They have their own experience, knowledge, skills, and approaches to learning. A student in a collaborative learning environment can take other students' views and ideas about the information provided in the learning material. The ideas coming out of the team can then be integrated as a deeper understanding of the material, or a better solution to a problem.

Traditional collaborative learning occurred in the form of face to face group discussion or problem solving. As the internet and learning technologies develop, online collaborative learning environments come out and are playing more and more important roles. For example, MOOCs (Massive Open Online Courses) have drawn massive number of learners. Learners in MOOCs are connected by the internet and can easily interact with each other using various types of tools, such as forums, blogs and social networks [23]. These digitized environments make it possible to track the learning processes in collaborative learning environments in greater detail.

Communication is one of the main factors that differentiates collaborative learning from individual learning [4; 6; 9]. As such, chats from collaborative learning environments provide rich data that contains information about the dynamics in a learning process. Understanding massive chat data from collaborative learning environments is interesting and challenging. Many tools have been invented and used in chat data analysis, such as LIWC (linguistic inquiry and word count) [12], Coh-Metrix [10], and topic modeling, just to name a few. Epistemic network analysis (ENA) has been playing a unique role in analyzing chat data from epistemic games [18]. ENA is rooted in a specific theory of learning: the epistemic frame theory, in which the collection of skill, knowledge, identity, value and epistemology (SKIVE) forms an epistemic frame. A critical theoretical assumption of ENA is that the connections between the elements of epistemic frames are critical for learning, not their presence in isolation. The online ENA toolkit allows users to analyze chat data by comparing the connections within the epistemic networks derived from chats. ENA visualization displays the clustering of learners and groups and the network connections of individual learners and groups. ENA requires coded data which has traditionally relied on hand coded data sets or classifiers that rely on regular expression mapping. Combining topic modeling with ENA will provide a new mode of preparing data sets for analysis using ENA.

In this study, we used a combination of topic modeling and ENA to analyze chat data to see if we could detect differences between the connections made by students with high learning gains versus students with low learning gains. Incorporating topic modeling

with ENA will make the analytic tool more fully automated and of greater use to the research community.

2. RELATED WORK

Chats have two obvious features. First, they appear in the form of text. Therefore, any text analysis tool may have a role in chat analysis. Second, chats come from individuals' interaction, which reflects social dynamics between participants. Therefore, a combination of text analysis and social network analysis should be helpful in understanding underlying chat dynamics. For instance, Tuulos et al. [21] combined topic modeling with social network analysis in chat data analysis. They found that topic modeling can help identify the receiver of chats (the person who a chat is given to).

In a similar effort, Scholand et al. [16] combined LIWC and social network analysis to form a method called "social language network analysis" (SLNA). The social networks were formed by counting the number of times chat occurred between any two participants. Based on the counts, participants were clustered into a tree structure, representing the level of subgroups the participants belong to. LIWC was then used to get the text features of chats. It was found that, some LIWC features were significantly different between in group conversations and out of group conversations.

Researchers have also recently explored the advantages of combining SNA (social network analysis) with deeper level computational linguistic tools, like Coh-Metrix. Coh-Metrix computes over 100 text features. The five most important Coh-Metrix features are: narrativity, syntax simplicity, word concreteness, referential cohesion and deep cohesion. Dowell and colleagues [8] explored the extent to which characteristics of discourse diagnostically reveals learners' performance and social position in MOOCs. They found that learners who performed significantly better engaged in more expository style discourse, with surface and deep level cohesive integration, abstract language, and simple syntactic structures. However, linguistic profiles of the centrally positioned learners differed from the high performers. Learners with a more significant and central position in their social network engaged using a more narrative style discourse with less overlap between words and ideas, simpler syntactic structures and abstract words. An increasing methodological contribution of this work highlights how automated linguistic analysis of student interactions can complement social network analysis (SNA) techniques by adding rich contextual information to the structural patterns of learner interactions.

In another study, Dowell et al. [7] showed that students' linguistic characteristics, namely higher degrees of narrativity and deep cohesion, are predictive of their learning. That is, students engaged in deep cohesive interactions performed better.

In the present research, we explore collaborative interaction chat data using the combination of topic modeling and epistemic network analysis. While previous studies focused on the relationship between language features and social network connections, our study focuses on prediction learning performance by semantic network connections students make in chats.

3. METHODS

Participants. Participants were enrolled in an introductory-level psychology course taught in the Fall semester of 2011 at a large university in the USA. While 854 students participated in this course, some minor data loss occurred after removing outliers and those who failed to complete the outcome measures. The final sample consisted of 844 students. Females made up 64.3% of this final sample. Within the population, 50.5% of the sample identified as Caucasian, 22.2% as Hispanic/Latino, 15.4% as Asian American, 4.4% as African American, and less than 1% identified as either Native American or Pacific Islander.

Course Details and Procedure. Students were told that they would be participating in an assignment that involved a collaborative discussion on personality disorders and taking quizzes. Students were told that their assignment was to log into an online educational platform specific to the University at a specified time, where they would take quizzes and interact via web chat with one to four random group members. Students were also instructed that, prior to logging onto the educational platform, they would have to read material on personality disorders. After logging into the system, students took a 10 item, multiple choice pretest quiz. This quiz asked students to apply their knowledge of personality disorders to various scenarios and to draw conclusions based on the nature of the disorders. The following is an example of the types of quiz questions students were exposed to:

- *Jacob was diagnosed with narcissistic personality disorder. Why might Dr. Simon think this was the wrong diagnosis?*
- *Dr. Level has measured and described his 10 mice of varying ages in terms of their length (cm) and weight (g). How might he describe them on these characteristics using a dimensional approach?*
- *Danielle checks her facebook page every hour. Does Danielle have narcissistic personality disorder?*

After completing the quiz, they were randomly assigned to other students who were waiting to engage in the chatroom portion of the task. When there were at least 2 students and no more than 5 students (M = 4.59), individuals were directed to an instant messaging platform that was built into the educational platform. The group chat began as soon as someone typed the first message and lasted for 20 minutes. The chat window closed automatically after 20 minutes, at which time students took a second 10 multiple-choice question quiz. Each student contributed 154.0 words on average (SD = 104.9) in 19.5 sentences (SD = 12.5). As a group, discussions were about 714.8 words long (SD = 235.7) and 90.6 sentences long (SD = 33.5).

An excerpt of a collaborative interaction chat in a chat room is shown below in Table 1. (student names have been changed):

Table 1. An excerpt of a collaborative interaction chat

Student	*Chat Text*
Art	ok cool, everyone's here. sooo first question
Art	ok so the certain characteristics to be considered to have a personality disorder?
Shaffer	Alright sooo first question: Based on these criteria describe several reasons why a psychologist might not label someone with grandiose thoughts as having narcissistic personality disorder?
Shaffer	hahaha never mind
Shaffer	that was the second question.
Art	lol its all good
Shaffer	okay so certain characteristics: doesn't it have to be like a stable thing?
Carl	i think the main thing about having a disorder is that its disruptive socially and/or makes the person a danger to himself or others

Vasile	yes, stable over time
Shaffer	yeah, and it also mentioned it can't be because of drugs
Art	also they have to have like unrealistic fantasies
Nia	yeah and not normal in their culture
Carl	no drugs or physical injury
Vasile	begins in early adulthood or adolescence
Shaffer	i think that covers them? haha
Art	ok, so arrogance doesn't just define it, they have to have most of these characteristics
Art	yeah i think we got them
Shaffer	is it most or is it like 6?

From the above excerpt, we can see several obvious things. First, the lengths of the utterances varied from one single word to multiple sentences. This needs to be considered in text analysis because some methods work only for longer texts. For example, Coh-Metrix usually works well for texts with more than 200 words. Topic modeling also needs enough length to reliably infer topic scores. Second, the number of utterances each participant gave were different. From how much and what a member said, we can see each member played a different role in that chat. Third, the ordered sequence of the utterances forms a time series. Understanding and visualizing the underlying discourse dynamics are important for meaning making with this type of data.

The data set contained 15,670 utterances, pretest scores (the first quiz) and post test scores (the second quiz) for 844 students, grouped in 182 chat rooms. Each chat room had 2 to 5 students, 4.73 by average. The average speech turns each student gave was 18.2 and the average speech turns in each room was 86.1.

The average pretest score was 36.01% correct and the average post-test scores 45.73% correct. Paired sample test shows that the post-test is significantly higher ($t= 14.13, N= 844$). We computed the learning gain of each student, using the formula

$$gain= \frac{\text{post} - \text{pre}}{1-\text{pre}}.$$

For all students ($N= 844$), the average learning gain is 0.11, 59.5% had positive learning gains above 0.1. 16.5% had the same scores and 23% had negative learning gains. Not surprisingly, students who had lower pretest scores had higher learning gains because they had greater potential to learn. Figure 1 shows the average learning gain as function of pretest score.

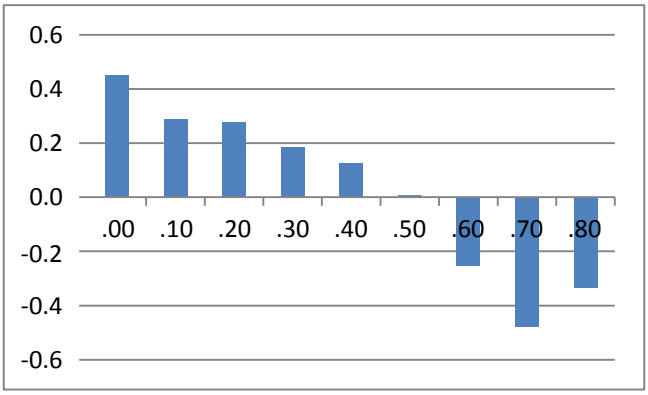

Figure 1. Average learning gain as a function of pretest score.

For students with pretest scores less than 50% correct (N=624), the average learning gain is 0.88, 69.7% had positive learning gains, 15.7% had the same scores and 14.6% had negative learning gains.

This data set has been analyzed in multiple studies. Cade et al. [3] analyzed the cohesion of the chats and found that deep cohesion of the chats predicts the students feeling of power and connectedness to the group. Dowell et al. [7] found that some Coh-Metrix measures predicts learning. Coh-Metrix measures describe common textual features that are not content specific. For example, cohesion is about how text segments are semantically linked to each other, which has nothing to do with what the text content is about. In this study, we use topic modeling to provide content dependent features and use epistemic network analysis to explore how the topics were associated in the chats.

4. TOPIC MODELING

Topic modeling has been widely used in text analysis to find what topics are in a text and what proportion/amount of each topic is contained. Latent Dirichlet Allocation (LDA) [2; 24] is one of the most popular methods for topic modeling. LDA uses a generative process to find topic representations. LDA starts from a large document set $D= \{d_1, d_2, \cdots, d_n\}$. A word list $W= \{w_1, w_2, \cdots, w_m\}$ is then extracted from the document set. LDA assumes that the document set contains a certain number of topics, say, K topics. Each document has a probability distribution over the K topics and each topic has a probability distribution over the given list of words. When a document was composed, each word that occurred in a document was assumed to be drawn based on the document-topic probability and the topic-word probability. For a given corpus (document set) and a given number of topics K, LDA can compute the topic assignment of each word in each document.

For a given topic, the word probability distribution can be easily computed from the number of times each word was assigned to the given topic. The beauty of topic modeling is that the "top words" (words with highest probabilities in a topic) usually give a meaningful interpretation of a topic. The distributions are the underlying representation of the topics. The top words are usually used to show what topics are contained in the corpus.

By counting the number of words assigned to each topic, a topic proportion score can be computed for each document on each topic. The topic proportion scores then become a document feature that can be used in further analysis. However, the proportion scores are based on the statistical topic assignment of words. When documents are very short, such as most utterances in our chat data, the topic proportion scores won't be reliable. Cai et al. [4] argued that alternative ways to compute document topic scores are possible.

4.1 TASA Topic Model

Although our chat data set contained 15,670 utterances, the utterances were short and the corpus is not large enough to build a reliable topic model. To get a reliable model, we used a well known corpus provided by TASA (Touchstone Applied Science Associates). This corpus contained documents on seven known categories, including business, health, home economics, industrial arts, language arts, science and social studies. Our content topic, personality disorders, is obviously in the health category. Of course, not all topics in TASA are relevant to our study. Therefore, after building up the model, we need to select relevant topics. We will cover that in the next sub-section.

There are a total of 37,651 documents in TASA corpus, each of which is about 250 words long. Before we ran LDA, we filtered out very high frequency words and very low frequency words. High frequency words, such as "the", "of", "in", etc., won't contain much topic information. Rare words won't contribute to meaningful statistics. 28,483 words (it might be better to say "terms") were left after filtering. A model with 300 topics was constructed by LDA.

4.2 Topic score computation and topic selection

From the TASA topic model, we computed the word-topic probabilities based on the number of times a word was assigned to each of the 300 topics. Thus, each word is represented by a 300 dimensional probability distribution vector. For each chat in our chat corpus, we simply summed up the word probability vectors for the words appeared in each chat. That gave us 300 topic scores for each chat. Recall that, the chats were associated with a reading material and two quizzes. While the students were free to talk about anything, the content of the reading material and the quizzes set up the main chat topics, that is, personality disorders.

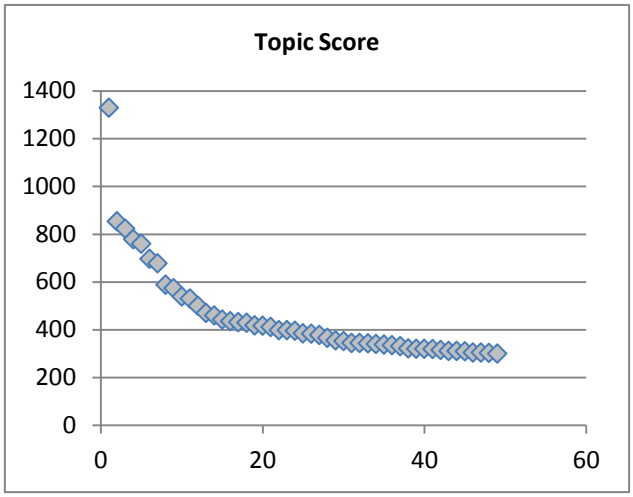

Figure 2. Sorted topic scores for topic selection.

The first thing we needed to do then was to investigate whether or not the "hot" topics from the computation made sense. To find that out, we computed the sum of all topic scores over all chats. The topics were sorted according the total topic score. The hottest topic had a total score higher than 1300, much higher than the second highest (less than 900). By examining the top words, this topic is about "illness", which is highly relevant to personality disorders. Six hot topics scored in the range from 600 to 900. They are about "outdoors", "biology", "people/social", "education" and "healthcare". The top words are listed below.

- **Illness:** health, disease, patient, body, diseases, medical, stress, mental, physical, heart, doctor, problems, cause, person, patients, exercise, illness, problem, nurse, healthy
- **Outdoors:** dog, energy, plants, earth, car, light, food, heat, words, animals, music, rock, language, children, air, uncle, city, sun, women, plant
- **Biology:** cells, cell, genes, chromosomes, traits, color, organisms, sex, egg, species, gene, body, male, female, parents, nucleus, eggs, sperm, organism, sexual
- **Psychology:** behavior, learning, theory, environment, feelings, sexual, physical, social, sex, human, research, person, animal, mental, response, positive, stress, personality, subject, reaction
- **People/Social:** joe, pete, mr, charlie, dad, frank, billy, tony, jerry, 'll, mom, 'd, going, 're, got, boys, looked, asked, paper, go
- **Education:** students, teacher, teachers, child, children, student, school, education, schools, learning, parents, tests, test, program, teaching, behavior, skills, reading, team, information
- **Healthcare:** patient, doctor, health, hospital, medical, dr, patients, nurse, disease, doctors, team, care, office, nursing, drugs, medicine, services, dental, diseases, help

"Illness", "biology", "psychology" and "healthcare" are the topics the learning materials involved. "Education" topic is about the education environment where the chat happened. "Outdoor" and "people/social" are off-task topics.

To get an idea about whether or not the topic scores were related to the learning gain, we aggregated the scores by person and computed the correlation between the total topic score and the learning gain for each topic. We were only interested in looking at the students with larger potential to learn, so we removed the data with pretest score greater than or equal to 0.5, leaving 624 students out of 844. The results (Table 1) showed that all topics were significantly correlated to learning gain. It doesn't seem to be great, because that seems to suggest that, whatever topic a student talked about, more a student talked, larger gain the student obtained. The real reason is that in the aggregation, all topic scores were summed up. Therefore, all topic scores were influenced by the chat length. So the correlation in Table 2 basically showed the chat length effect.

Table 2. Correlation between total topic scores and learning gain (N=624, pretest<0.5)

Topic	*Post-test*	*Pretest*	*Gain*
Illness	.183**	.116**	.132**
Outdoors	.216**	.133**	.154**
Biology	.159**	.125**	.105**
Psychology	.182**	.096*	.140**
People/Social	.115**	.022	.107**
Education	.175**	.118**	.121**
Healthcare	.157**	.130**	.097*

To remove the chat length effect, the simplest way is to divide all scores by the number of words (terms) in each chat. However, in this study, to be consistent with subsequent analysis, we normalized the topic scores to topic proportion scores by dividing each topic score for each utterance by the sum of all seven topic scores of the same utterance.

The results (Table 3) showed that the topic "people/social" had a significant negative correlation to learning gain. Others were not significant but were in the direction we would expect. "Illness", "biology", "psychology" and "healthcare" were positively correlated with gain scores, while "outdoors" and "people/social" topics were negatively correlated with gains scores. We observed almost no correlation for the "Education" topic. This seems to indicate that the aggregated topic scores have limited power in predicting learning. Therefore, we used ENA to examine the connections or association of these topics in the students discourse to

develop a predictive model of learning gains based on the use of these topics.

Table 3. Correlation between normalized topic proportion scores and learning gain (N=624, pretest<0.5)

Topic	*Post-test*	*Pretest*	*Gain*
Illness	.099*	0.077	0.067
Outdoors	-0.063	-0.043	-0.044
Biology	.085*	0.054	0.063
Psychology	0.067	0.019	0.058
People/Social	-.127**	-0.076	-.083*
Education	0.027	0.056	-0.002
Healthcare	0.073	.096*	0.027

5. EPISTEMIC NETWORK ANALYSIS

ENA measures the connections between elements in data and represents them in dynamic network models. ENA creates these network models in a metric space that enables the comparison of networks in terms of (a) difference graph that highlights how the weighted connections of one network differ from another; and (b) statistics that summarize the weighted structure of network connections, enabling comparisons of many networks at once.

ENA was originally developed to model cognitive networks involved in complex thinking. These cognitive networks represent associations between knowledge, skills, habits of mind of individual learners or groups of learners. In this study, we used ENA to construct network models. For each individual student, we constructed an ENA network using the selected seven topic scores for each utterance the student contributed to the group.

5.1 Process

While the process of creating ENA models is described in more detail elsewhere (e.g. [11; 17-19]), we will briefly describe how ENA models are created based on topic modeling. Here we defined network nodes as the seven topics identified from the topic model. We defined the connections between nodes, or edges, as the strength of the co-occurrence of topics within a moving stanza window (MSW) of size 5 [19]. To model connections between topics we used the products of the topic scores summed across all chats in the MSW. That is, for each topic, the topic scores are summed across all 5 chats in the MSW. Then ENA computed the product of the summed topic loadings for each pair topics to measure the strength of their co-occurrence. For example, if the sum of the topics scores across five chats was 0.5 for "illness", 0.3 for "psychology", and 0.2 for "healthcare", these scores would result in three co-occurrences, "illness-psychology", "illness-healthcare", and "psychology-healthcare", with scores of 0.15, 0.1, and 0.06, respectively.

Next ENA created adjacency matrices for each student that quantified the co-occurrences of topics within the students' discourse in the context of their chat group. Subsequently, the adjacency matrices were then treated as vectors in a high dimensional space, where each dimension corresponds to co-occurrence of a pair of topics. The vectors were then normalized to unit vectors. Notice that the normalization removed the effect of chat length embedded in the topic scores. A singular value decomposition (SVD) was then performed for dimensional reduction. ENA then projected a vector for each student into a low dimensional space that maximizes the variance explained in the data. Finally, the nodes of the networks, which in this case correspond to the seven selected topics generated from TASA corpus, were placed in the low dimensional space. The topic nodes were placed using an optimization algorithm such that the overall distances between centroids (centers of the mass of the networks) and the corresponding projected student locations was minimized. A critical feature of ENA is that these node placements are fixed, that is, the nodes of each network are in the same place for all units in the analysis. This fixing of the location of the nodes allows for meaningful comparisons between networks in terms of their connection patterns which allow us to interpret the metric space. As a result, ENA produced two coordinated representations: (1) the location of each student in a projected metric space, in which all units of analysis included in the model were located, and (2) weighted network graphs for each student, which explained why the student was positioned where it was in the space.

ENA also allows us to compare the mean network graphs and mean position in ENA space between different groups of students. In this study, we only considered the students with high potential to learn, i.e., the 624 students with pretest score < 0.5 (50% correct). Among these students, we compared the networks of low learning gain students (gain<-0.1, N=194) with the networks of high learning gain students (gain>0.43, N=105). We compared these groups using difference network graph, which was formed by subtracting the edge weights of the mean discourse network for the low gain group students from the mean discourse network from the high gain group. This difference network graph shows us which topic connections are stronger for each group. In addition, we conducted a *t*-test to test the difference between group means.

5.2 Results

Figure 3 shows mean discourse networks for students with low gain scores (left, red), students with high gain scores (right, blue), and a difference network graph (center) that shows how the discourse patterns of each group differs. Students with low gains had stronger connections between the "people/social" topic and all other topics except for "illness". More importantly, the connection that was the strongest for low gain students compared to high gain students was between "people/social" and "outdoors". Students with high gain scores made stronger connections between the topics of "illness", "psychology", "healthcare", "biology", and "education".

Table 4. Comparison of centroids between low gain and high gain students, p=0.04, t= 2.0, SD

	N	Mean	SD
High gain	105	0.033	0.220
Low gain	194	-0.048	0.322

Figure 4 shows centroids, or the centers of mass, of individual students' discourse networks and their means with low gain score students in red and high gain score students in blue. The differences between these two groups were significant on the x dimensions (see table 4). This means that the differences we saw in figure 2 and described above are statistically significant. In other words, the high learning gain students' discourse was more towards the right side of the ENA space and the low learning gain students' discourse was more towards the left side. That indicates that the discourse of students with high learning gains made more connections between on-task topics ("illness", "psychology", "healthcare", "biology", and "education"), while the discourse of

low gain students made more connections between off-task topics ("people/social" and "outdoors").

6. DISCUSSION

ENA makes it possible to visualize the chat dynamics to help researchers gain deeper understanding of what is going on in a collaborative learning environment. Differences in what topics students connect in discourse can predict learning outcomes. Previous use of ENA has relied on human coded data or use of regular expressions to classify data. Utilizing topic modeling can lead to fully automated ENA, making it more accessible to a wider group of researchers and allows ENA to be used with more and larger data sets.

The fact that the epistemic network predicts learning validates further application of ENA. For example, the turn by turn chat dynamics can be plotted as trajectories in the 2-D space, where the topics are placed. Investigating the trajectory patterns and their relationship to learning or socio-affective components are interesting future research directions.

We used a general topic model in this study. Many studies in the literature used LDA for topic modeling on relatively small corpora. This causes two problems. 1) LDA topic models built upon small corpora are not reliable, because LDA requires large number documents with relatively large size for each document. Inadequate corpus can result in misleading results. 2) Using a topic model that is not common would result in arbitrary interpretation. For example, the representation of "illness" from different corpus could be very different. Therefore, it is hard to compare the claims made to "illness" across different studies. Using a reliable, common topic models will set up a common language for different studies.

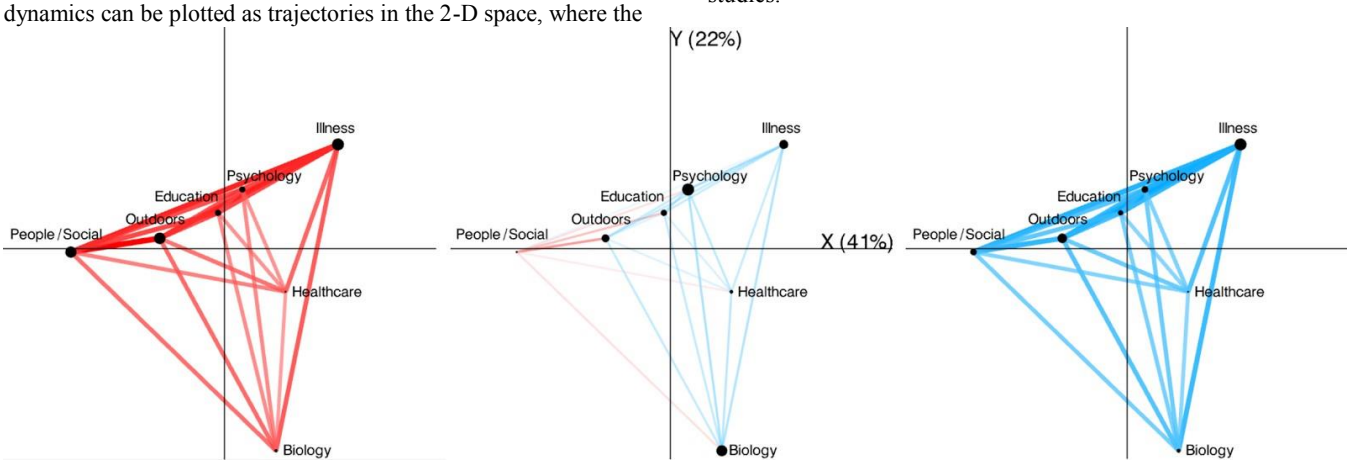

Figure 3: Mean discourse networks for students with low gain scores (left, red), students with high gain scores (right, blue), and a difference network graph (center).

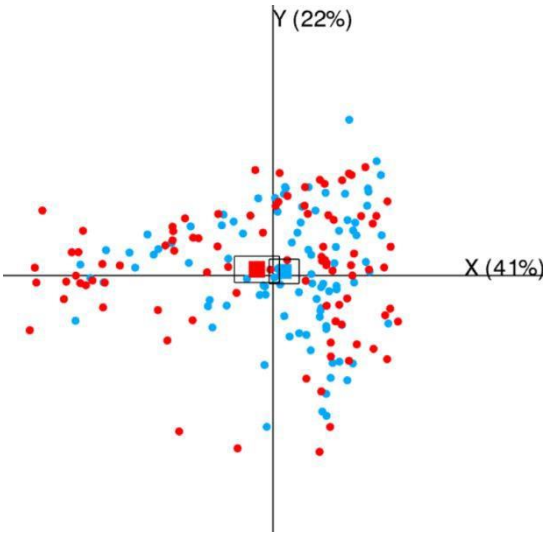

Figure 4: Discourse network centroids low gain score students red, high gain score students blue.

Topic scores for documents are usually inferred from topic models. While for longer documents, the topic scores can be used in many applications (e.g., text clustering [1]), the inferred topic proportion scores won't be useful for analyzing chats if we need to treat each utterance as a unit of analysis. It is not useful because chat utterances are too short. The statistical inference algorithm contains a high degree of randomness for short documents. As an extreme example, an utterance with a single word, would result in inferred topic proportion scores with "1" on one topic and "0" on others. The problem is that, this "1" was assigned to a topic with certain degree of uncertainty. That is, the topic this "1" was assigned to could be any topic. While aggregated analysis may not be sensitive to such uncertainty, detailed utterance by utterance analysis would suffer from it.

Our method of computing topic scores is based on the topic probability distribution over each word. We treat the topic distribution of each word as a vector. When computing the topic score, the simple sum of all word vectors gives scores to all topics. As we have pointed out, the summation algorithm will have a length effect. Therefore, when such topic scores are used, removing length effects through normalization is necessary. In this article, we did not use weighted sum as suggested in Cai et al. [4]. Comparing the effect of different weighting is beyond the scope of this paper.

When a general topic model is used, selecting topics relevant to the specific analysis becomes important. Our approach was to look at the total scores of utterances and find the "hot" topics by sorting the total topic scores. In our study, we had a quickly decreasing curve that helped us to select topics. We believe this would be the case for most studies using a model containing far more topics than the topics contained in the target data.

Although our study started with topic modeling to capture the "what" in the chats, the association networks constructed in the epistemic network analysis actually turned the "what" into a "how": how the topics in the chats associated with each other. This is conceptually similar to the cohesion features Dowell [7] and Cade [3] used.

Topic modeling emphasizes content words. When a topic model is built, stop words are usually removed. An interesting question is, what if we do the opposite: keep stop words and remove content words? Pennebaker (e.g., [13]) laid foundational work in this direction. The LIWC tool Pennebaker and his colleagues created provides over a hundred text measures by counting non-content words. LIWC measures could provide different features to epistemic network analysis and reveal different aspects of the chat dynamics.

7. ACKNOWLEDGMENTS

The research on was supported by the National Science Foundation (DRK-12-0918409, DRK-12 1418288), the Institute of Education Sciences (R305C120001), Army Research Lab (W911INF-12-2-0030), and the Office of Naval Research (N00014-12-C-0643; N00014-16-C-3027). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, or DoD. The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of researchers from psychology, computer science, and other departments at University of Memphis (visit http://www.autotutor.org).

8. REFERENCES

[1] Alghamdi, R. and Alfalqi, K. 2015. A Survey of Topic Modeling in Text Mining. *IJACSA) International Journal of Advanced Computer Science and Applications*. 6, 1 (2015), 147–153.

[2] Blei, D.M., Edu, B.B., Ng, A.Y., Edu, A.S., Jordan, M.I. and Edu, J.B. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 3, (2003), 993–1022.

[3] Cade, W.L., Dowell, N.M.M. and Pennebaker, J. 2014. Modeling Student Socioaffective Responses to Group Interactions in a Collaborative Online Chat Environment. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*. 2, 21 (2014), 399–400.

[4] Cai, Z., Li, H., Graesser, A.C. and Hu, X. 2016. Can Word Probabilities from LDA be Simply Added up to Represent Documents? *Proceedings of the 9th International Conference on Educational Data Mining*. (2016), 577–578.

[5] von Davier, A.A. and Halpin, P.F. 2013. Collaborative Problem-Solving and the Assessment of Cognitive Skills: Psychometric Considerations. *ETS Research Report Series*. December (2013), 36 p.

[6] Dillenbourg, P. and Traum, D. 2006. Sharing Solutions: Persistence and Grounding in Multimodal Collaborative Problem Solving. *The Journal of the Learning Sciences*. 15, 1 (2006), 121–151.

[7] Dowell, N., Cade, W., Tausczik, Y., Pennebaker, J., and Graesser, A. 2014. What Works: Creating Adaptive and Intelligent Systems for Collaborative Learning Support. *Springer International Publishing Switzerland*. (2014), 124–133.

[8] Dowell, N.M.M., Skrypnyk, S., Joksimović, S., Graesser, A., Dawson, S., Gašević, D., Hennis, T. a., Vries, P. De and Kovanović, V. 2015. Modeling Learners ' Social Centrality and Performance through Language and Discourse. *Educational Data Mining - EDM'15* (2015), 250–257.

[9] Fiore, S.M., Rosen, M. a., Smith-Jentsch, K. a., Salas, E., Letsky, M. and Warner, N. 2010. Toward an understanding of macrocognition in teams: predicting processes in complex collaborative contexts. *Human factors*. 52, 2 (2010), 203–224.

[10] Graesser, A.C., McNamara, D.S., Louwerse, M.M. and Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*. 36, 2 (2004), 193–202.

[11] Li, H., Samei, B., Olney, A., Graesser, A. and Shaffer, D. 2014. Question Classification in an Epistemic Game. *International Conference on Intelligent Tutoring Systems*. (2014).

[12] Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. 2015. The Development and Psychometric Properties of LIWC2015. *Austin, TX: University of Texas at Austin*. (2015).

[13] Pennebaker, J.W., Chung, C.K., Frazee, J. and Lavergne, G.M. 2014. When Small Words Foretell Academic Success : The Case of College Admissions Essays. (2014), 1–10.

[14] Rosen, Y. 2014. Assessing Collaborative Problem Solving Through Computer Agent Technologies. *Encyclopedia of information science and technology*. 9, November (2014), 94–102.

[15] Sawyer, R.K. 2014. The new science of learning. *The Cambridge Handbook of the Learning Sciences*. 1–18.

[16] Scholand, A.J., Tausczik, Y.R. and Pennebaker, J.W. 2010. Assessing group interaction with social language network analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 6007 LNCS, (2010), 248–255.

[17] Shaffer, D.W. 2006. Epistemic frames for epistemic games. *Computers and Education*. 46, 3 (2006), 223–234.

[18] Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A.A. and Mislevy, R.J. 2009. Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*. 1, 2 (2009), 33–53.

[19] Siebert-Evenstone, A.L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A.R. and Shaffer, D.W. 2016. In search of conversational grain size: Modeling semantic structure using moving stanza windows. *International Conference of the Learning Sciences*. (2016).

[20] Slavin, R.E. 1995. Cooperative Learning: Theory, Research and Practice (2nd Ed.). *The Nature of Learning*. (1995), 208.

[21] Tuulos, V.H. and Tirri, H. 2004. Combining Topic Models and Social Networks for Chat Data Mining. *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*. October (2004), 206–

213.

[22] Whitepaper, A.R. 2014. What Happens When We Learn Together. (2014).

[23] Yousef, A.M.F., Chatti, M.A., Schroeder, U., Wosnitza, M. and Jakobs, H. 2014. A Review of the State-of-the-Art. *Proceedings of the 6th International Conference on Computer Supported Education - CSEDU2014*. (2014), 9–20.

[24] Wang Z., Qiu B., Bai, W., Chuan, S. and Le, Y. 2014. Collapsed Gibbs Sampling for Latent Dirichlet Allocation on Spark. *JMLR: Workshop and Conference Proceedings*. 2004 (2014), 17–28.