Epistemic Network Analysis of students' longer written assignments as formative/summative evaluation

Simon Skov Fougt
Metropolitan University College
Nyelandsvej 27-29, 2000 Frederiksberg
Denmark
sifo@phmetropol.dk

Amanda Siebert-Evenstone
University of Wisconsin-Madison
Educational Sciences Bldg., 1025 W.
Johnson St., Madison, WI
USA
alevenstone@wisc.edu

Brendan Eagan
University of Wisconsin-Madison
Educational Sciences Bldg., 1025 W.
Johnson St., Madison, WI
USA
eaganb@gmail.com

Sara Tabatabai
University of Wisconsin-Madison
Educational Sciences Bldg., 1025 W.
Johnson St., Madison, WI
USA
tabatabai2@wisc.edu

Morten Misfeldt
University of Aalborg
A.C. Meyers Vænge 15, 2450
Copenhagen SV
Denmark
misfeldt@learning.aau.dk

ABSTRACT

This paper reports on an exploratory trial of developing pedagogical visualizations of 16 students' written assignments on literary analysis using two sets of keywords and Epistemic Network Analysis (ENA). The visualizations are aimed at summative evaluation as a tool for the professor to support assessment and understanding of subject learning. Results show that ENA can visually distinguish low, middle and high performing students, but not statistically significantly. Thus, our trial provides a tool for the professor that supports understanding of subject learning and formative assessment.

CCS CONCEPTS

• **Human-centered computing~Visualization toolkits**
• *Information systems~Data analytics*

KEYWORDS

Epistemic Network Analysis, pedagogical learning analytics, formative and summative assessment, student assignment

ACM Reference format:

S.S. Fougt, A. Siebert-Evenstone, B. Eagan, S. Tabatabai, and M. Misfeldt. 2018. Epistemic Network Analysis of students' longer written assignments as formative/summative evaluation. In *Proceedings of the International Conference on Learning Analytics and Knowledge, Sydney, Australia, March 2018 (LAK'18)*, 5 pages. https://doi.org/10.1145/3170358.3170414

LAK '18, March 7–9, 2018, Sydney, NSW, Australia

ACM 978-1-4503-6400-3/18/03…$15.00
https://doi.org/10.1145/3170358.3170414

1 INTRODUCTION WITH RESEARCH QUESTIONS

This paper reports on an exploratory trial from The Metropolitan Teachers College, Denmark, where Epistemic Network Analysis (ENA) [14, 15] was used to develop pedagogical visualizations of the discursive network of relevant subject terms of 16 teacher students' longer written assignments on literacy analysis, aimed as a tool for the professor to support assessment and understanding of subject learning. Our future intention is to develop ENA into an independent formative assessment tool for instructors, and this trial investigates potentials and pitfalls.

Digital support for assessment of students' writing can benefit student learning, save time and resources and improve the quality of educational experiences [16]. But such development is potentially problematic since it could lead to redundant and simplistic practices [7]. In this paper, we will address the following research questions:

- How, and to what extent, can ENA predict the quality of longer student assignments?
- What characterizes the epistemic network of low, middle, and high performers?

After a brief presentation of approaches to research in automatic discourse analysis and pedagogical use, we present the case and elaborate on our ENA method. Then we present examples of analysis followed by a discussion and a conclusion.

2 APPROACHES TO TEXT ANALYSIS

Many tools have been created to analyze text data, e.g. LIWC [linguistic inquiry and word count; 2], Coh-Metrix [10], topic modeling [2], and others. For example, Crossley et al. [6] have utilized natural language processing (NLP) tools to assess written responses. Cai et. al. [4] combined topic modeling and ENA to analyze chat data and assess the differences between low gain and

 S.S. Fougt et al.

high gain students in a collaborative learning environment. These tools are used to classify large amounts of text for subsequent analysis and have been successful at estimating cognitive, motivational, and relational constructs. However, they face the disadvantage of weaker predictive accuracy across contexts [11].

While there are many methods of learning analytics, there is a danger in using statistics without theory. Wise and Shaffer [17] argue that theory can provide guidance for variable choices and potential confounds, provide a framework for interpretation, and generalize results to other contexts. Therefore, we used a theory-based approach to assess the quality of students' assignments.

3 THEORY

Several learning theories describe complex thinking as understanding connections among domain elements. Chi and colleagues [5] showed that experts organize their physics knowledge and understandings in systematically different ways than novice physics students. Novices are more likely to group physics problems by surface features, while experts grouped problems based on the underlying principles of that problem. Bransford and colleagues [3] argued that expert thinking cannot be reduced to knowing a large number of isolated facts. Expertise is characterized by understanding which ideas are applicable in specific contexts by having a rich understanding of meaningful patterns of information. Extending these ideas, Shaffer [13] characterizes learning as developing an epistemic frame, which is a pattern of associations among domain elements, habits of mind, and cognitive elements that characterize a community of practice.

Therefore, a good model of student thinking needs to be able to analyze the relationships among elements in the domain. ENA is one such tool, which analyzes the structure of connections in student discourse by looking at the co-occurrence of concepts within the conversations, topics or activities that take place during learning. ENA creates a discourse network model of thinking by identifying the co-occurrence of skills, knowledge, values, and other elements of work in a particular community of practice [14]. The co-occurrences are identified within collections of related utterances, which are nested within conversations, a fundamental unit of analysis in ENA. In this study, we identify the paragraph in the students' assignments as the conversation because we assume that all ideas within a paragraph are related to one another, while ideas across paragraphs do not.

In order to analyze the data, researchers and teachers need a way to identify and describe key ideas in a given set of data, sometimes referred to as coding [9]. A common way of identifying these codes is through grounded theory, where researchers inductively identify concepts within the text which are then grounded in that text [8]. This is in contrast to deductive codes which are based on theory or inclinations of the researcher. For ENA to be useful for teachers, teacher must create codes and use skills typically associated with education researchers.

In this paper, we work to bridge the gap between teachers and researchers by building on the existing work that teachers already do as part of their classroom management and student assessment. Of course, teachers engage in formative and summative assessments, but they also group students by performance levels, test the effectiveness of new activities, and compare results against past experiences.

Arastoopour argued that effective assessments use components of Evidence-Centered Design assessment which use evidence from student work to evaluate learning and understanding [1]. Similarly, this paper attempts a new method for identifying better assessments of student learning.

Once a set of data has been coded, ENA is able to create models based on minimal inputs for making it easy to use for non-researchers. In this paper, we investigate automated classification with modeling techniques as a cutting-edge way for teachers to understand and analyze complex thinking in student writing.

4 METHOD

4.1 Case

Danish Teacher Education lasts four years where students study pedagogical subjects and additionally choose three teaching subjects, where one has to be Danish (L1) or Mathematics. Danish consists of four modules over two years: *Language*, *Fictional texts*, *Literacy* and *Multimodal texts*. The present study was undertaken in the Fictional texts-module during second semester, where students are trained to teach fictional texts in grade 4-10 (age 10-16). The students engage in 66 lessons of teaching, and the module is passed by writing an approved five-page essay. In this case, students analyze a fictional text by using either a biographical, literary criticism, reader response, or phenomenological approach and then relate their analysis to teaching.

16 students completed the assignment and were evaluated traditionally by their professor, giving them a grade (A-F), and the empirical data thus consists of these 16 essays. The sample contained one biographical analysis, four literary criticisms, five reader responses, and six phenomenological analysis.

4.2 Coding student essays

To explore two approaches to modeling essays with ENA, we generated two sets of eight keywords. This first set consisted of eight general terms associated with literary analysis. We also had the professor generate eight specific keywords, four pairs each associated with the four specific literary analysis approaches in order to verify our analysis.

To turn each keyword into a useable automated code, we identified the stem of each word to find different versions and conjugations of that word. For instance, we searched for the equivalent of the word "interpret" to find instances of "interpret", "interpretation", "interpreted," and other versions of the key idea of interpretation. We applied this set of keywords to each paragraph of a student essay using regular expression matching.

We analyzed student essays using the following 16 keywords or disciplinary terms (codes) - the first eight are general literary analysis terms, the last eight are in pairs particularly linked with one of the analysis approaches:

General literary analysis terms:

1. *Report*: Explanation of the literary analysis theory
2. *Environment*: Setting in which the story takes part
3. *Narrator:* A fictive person, not the writer
4. *Motif-theme:* A recurring image or idea that works to unify or connect a piece using a common concept
5. *Interpretation:* The writer's understanding and meaning-making about a text
6. *Genre:* Established categories of writing
7. *Analysis:* Focused on components, themes and techniques
8. *Description:* A brief resume of the text, allowing the reader to understand the context and activities of the story

Literary analysis specific keywords:

9. *Authorship*: Linked to a biographical approach, where the authorship is expected mirrored in the text
10. *Mirror*: Linked to a biographical approach, where the authorships is expected mirrored in the text
11. *Strata*: Linked to a literary criticism approach, where the text is considered a system of layers or stratas
12. *Close reading*: Linked to a literary criticism approach, studying the interaction between the stratas in the texts
13. *Blanks*: Linked with a reader response approach, where the reader has to add content or understanding
14. *Ambiguous*: Linked with a reader response approach, texts can be interpreted in diverse ways
15. *Overdetermined:* Linked with a phenomenological analysis. Cultural patterns in texts that limits possible interpretation of blanks in texts
16. *Underdetermined*: Linked with a phenomenological analysis. Overdetermination limits possible interpretations of underdetermination.

4.3 Epistemic Network Analysis

Epistemic network analysis (ENA) measures the connections between coded elements in data and represents them in dynamic network models [14]. The technical details have been provided elsewhere [15], but briefly, ENA models the structure of connections among coded elements by quantifying the co-occurrences of codes within a defined segment of data, or stanza.

After defining the segmentation structure, ENA creates an adjacency matrix representing the co-occurrences of codes in each stanza. To construct an adjacency matrix, ENA assigns a one for each unique pair of codes that co-occur one or more times in those utterances, and a zero for each unique pair that does not co-occur in the stanza. ENA sums the adjacency matrices into a cumulative adjacency matrix, where each cell represents the number of stanzas (i.e., the number of adjacency matrices) in which that unique pair of codes was present. Each collection of co-occurrences is thus represented by a cumulative adjacency matrix that summarizes the pattern of connections among codes.

ENA then converts the cumulative adjacency matrices into cumulative adjacency vectors that are projected into a high-dimensional space based on the co-occurrence of codes across segments. These cumulative adjacency vectors are normalized to control for the length of response so that papers that write more words are not weighted more heavily than papers that had fewer words but make the same connections in their writing. ENA then performs a singular value decomposition on the normalized vectors. This produces a rotation of the original high-dimensional space, such that the rotated space provides a reduced number of dimensions that capture the maximum variance in the data.

The resulting models can be visualized as networks in which the nodes in the model are the codes and the lines connecting the nodes represent the co-occurrence of two codes. Thus, we can quantify and visualize the structure of connections among the keyword codes, making it possible to characterize student connection making patterns in their essays. ENA creates these network models in a metric space that enables the comparison of networks in terms of (a) complexity in terms of numbers of, and strengths of, connections; and (b) statistics that summarize the weighted structure of network connections, enabling comparisons of many networks at once.

In this study, we segmented essays into units consisting of paragraphs (e.g. theory, method), so from an ENA perspective, we had more units of analyses (n=469) than the number of student essays (n=16) This allowed us to model the connections student make throughout their papers. After measuring the co-occurrences within the sentence, ENA sums the matrices across all sentences in that paragraph, and all paragraphs in the paper.

4.4 Analysis

To analyze the essays we compared them based on performance level, which divided the essays into low, medium, and high performing papers based on their grades with the aim of being able to visually identify differences. We created models for essays using (1) eight general keywords, and (2) using eight general keywords and eight customized keywords. Both models bookend a range of the number keywords, or nodes, to consider using when modeling the complexity of student essays with ENA.

5 RESULTS

To investigate which connections accounted for differences between the three groups (low, medium, high), we compared their mean epistemic networks. In particular, groups with higher quality papers had more connections or more complex networks than lower quality papers. Thicker lines on the figures to come indicate stronger and more frequent connections.

Specifically, in our first analysis we created mean keyword epistemic networks for low (green), medium (blue), and high (red) performing groups of students, see Fig. 1. The connections distinguishing low from high scoring groups are connections to the *description* keyword. On average, the students that generated medium and high papers shows more connections between *description* and other elements of literary analysis than the papers for students that received a low grade on the assignment.

Overall, the networks with eight keywords indicate that higher performing students make more and different connections compared to middle and low performing students. The high performing group's mean network has 20 types of keyword connections, the medium performing group's mean network has 13 types of keyword connections, and the low performing group's

 S.S. Fougt et al.

mean network contains 10 types of connections. Beyond the simple measure of complexity given by the number of connection types, we statistically compared the means of the plotted points for each group which correlate to centers of mass for each mean keyword networks. We identified statistically significant differences on the y-dimension between low ($M=0.029$, $SD=0.107$) and middle ($M=-0.015$, $SD=0.214$) performing students, $t(282.99)=-2.328$, $p<0.05$.

The differences in connections between *description* and *narrator* as well as those between *description* and *interpretation* are driving the mean for the middle performing group lower on the y-dimension compared to the low performing group. The connections between *description* and *narrator* are stronger in the middle performing group and connections between *description* and *interpretation* are present in the middle performing group's keyword network, but absent in the low performing group network.

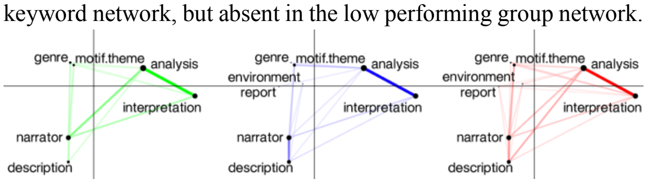

Figure 1: Mean epistemic networks of low (green, left), medium (blue, center), and high papers (red, right) representing the structure of connections across 8 keywords for each set of students' papers.

As mentioned, we repeated the same analysis, this time including the additional eight customized concepts to the ENA models, see **Fig. 2**. The original eight keywords were general literary analysis terms, whereas the eight additional literary analysis keywords consist of four pairs each linked with one of the literary approaches. Again, we visually identify a more complex network and different connections among higher performing students compared to middle and low performing students. In addition, middle performing students had a more complex network with connections not present in the low performing students' keyword network. We identified a statistically significant difference on the y-dimension between low ($M=-0.028$ $SD=0.107$) and middle ($M=0.014$, $SD=0.187$) performing students, $t(266.7)=2.419$, $p=0.016$.

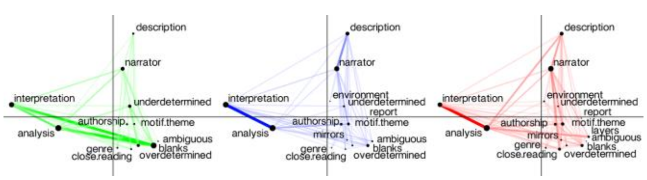

Figure 2: Mean epistemic networks of low (green, left), medium (blue, center), and high papers (red, right) representing the structure of connections across 16 keywords for each set of students' papers.

Specifically, we see that low performers most make connections between *blanks*, *interpretation*, and *analysis*. In contrast, the middle and high performers display networks with more balanced connections while still emphasizing the connection between interpretation and analysis. Both analysis matched with grade (low-medium-high), given by their professor.

6 DISCUSSION

The focus of this study was to explore whether ENA could provide a useful pedagogical tool for teachers to support their assessment of longer student writing assignments and to visualize understanding of subject learning. That is, we have demonstrated one way in which ENA could be used as an aid or assistant for instructors assessing students written assignments by allowing them to visualize and quantify the connections students make in their writing. Instructors can then tailor feedback or subsequent instruction of students based on their understanding of the connections students have made. The low performing networks in general show simpler network structures which make connections between fewer of the available keywords. The high performers show broader, more complex network structures making connections between all of the keywords. The middle performers lie between the low and high performers. The connections from and between *interpretation* and *analysis* are often stronger and more varied as performance improves.

Visually, ENA provides teachers and researchers with a tool that supports the teacher's original assessment, as was the case in this study, and provides an opportunity to interpret content learning through a new lens. ENA has the potential to indicate the quality of the student assignment, although we are generally not able to identify significant differences comparing low, middle and high performing students. And perhaps this a good result, as one of the possible abuses of the suggested use of ENA here would be to replace the teacher's assessment of student work. This was never our intention - the intention is to develop a tool that supports teacher's assessments and visualizes the ways that students construct written arguments.

One struggle with text analysis is choosing the right number and relevant concepts for keyword matching. In this case, the professor was instructed to choose five to eight keywords, based on our experience from former work with ENA that this amount of keywords adequately differentiates networks while maintaining an interpretable network graph. Our comparison of the two sets of analysis, using either eight or 16 keywords, provided the ability to conduct a general comparison as well as a deeper analysis taking into consideration the different types of literary analyses undertaken by the students. The use of 16 keywords created more nuanced networks which gave greater detail of the composition of the papers for the three performance levels. However, the increase in the number of keywords may come at the cost of noisy, harder to interpret networks.

Our two analyses start to articulate two dimensions of using ENA to analyze the complexity of student essays. The first is the number of codes, in our case keywords, to include in an ENA model. Researchers and instructors using ENA need five to eight codes, on the low end, in order to capture the complexity they are interested in modeling. On the high end, including too many codes can result in networks that are too complex to interpret or explain. If our second analysis had contained 32 nodes instead of 16, the networks would look like spaghetti and would be impossible for a researcher to make meaning in such a space. In addition, networks with too many codes can produce over-fit models.

A second dimension of understanding texts with ENA is understanding the benefits of domain-general keywords versus specific and customized keywords generated by the instructor. Our first analysis was based on general keywords and provided a useful model for assessing essays, but instructors, even those relying on standards, often want to customize their curriculum and assessments. Instructors using ENA to assess essays need to make decisions about which codes to use and what ratio of general versus customized codes to use.

More generally, keywords have to be aligned with a particular assignment, as well as present in the students' papers. This trial had a small sample size (n=16), and also therefore, our results are purely an indication of what might be possible. In this particular case, the students did not know the keywords, prior to writing. One risk of letting students know the keywords before writing would be students gaming the system and entering words to produce complex networks. This would of course be discovered by the teacher. On the other hand, the keywords are prototypical, and therefore students could probably predict some of them.

ENA provides a potentially helpful tool in combination with automated text analyses, where teachers themselves can list the accurate keywords based on their particular assignments. We are working to automate keyword creation by having teachers generate a name, definition and example answers and then extracting keyword strings to automate text analysis for teachers. This tool would facilitate teacher generation of a codebook which would provide easier access to learning analytic tools.

This method is obviously vulnerable to the use of non-disciplinary terms, and to some degree also to students' spelling. Issues with spelling and conjugation can be reduced by identifying the stem of each keyword, whereas the use of non-disciplinary terms is difficult to address. This again supports the need for a teacher.

In summary, this use of ENA supports the teacher when indicative and useful keywords are used, and in this way, ENA supports the teacher's understanding of subject learning. ENA so to speak visualizes, as stated in the theory section with Bransford and colleagues [3] and Shaffer [13], that patterns are key in complex thinking. And therefore, our model analyzes relationships in subject content concepts, and we are visually able to identify differences between low, middle and high performing students - but not significantly. In that the sense, our trial supports the need for a teacher to do a closer assessment of the papers, although ENA seems able to support it.

7 CONCLUSIONS

In this short paper, we have discussed the use of ENA on longer written assignments as a formative and summative evaluation tool for teachers based on a small trial of 16 student-teachers' written assignments. Our analyses indicate that ENA can visually confirm the quality of the assignments and analyze their complexity and provides the teacher with a useful tool.

Our first analysis with eight keywords was confirmed by the 16 keywords analysis. It could both be argued that only the simple analysis is needed to predict the quality. But using ENA as automated discourse analysis is an emerging field. Further studies could focus on choosing the right (number of) keywords.

In this case, keywords were provided, tested and discussed by experienced ENA users, and therefore future research is also needed on the teachers ability to choose the right keywords.

ACKNOWLEDGMENTS

This work was partly funded by the National Science Foundation (DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DRL-1661036, DRL-1713110, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, REC-0347000), the MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

REFERENCES

[1] G. Arastoopour, W. Collier, N.C. Chesler, J. Linderoth, and D.W. Shaffer. 2015. Measuring the complexity of simulated engineering design problems. *American Society for Engineering Education*.
[2] D.M. Blei, A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3, January: 993–1022.
[3] J.D. Bransford, A.L. Brown, and R.R. Cocking. 1999. *How people learn: Brain, mind, experience, and school*. National Academies Press, Washington, DC.
[4] Z. Cai, B. Eagan, N.M. Dowell, J.W. Pennebaker, D.W. Shaffer, and A. Graesser. 2017. Epistemic Network Analysis and Topic Modeling for Chat Data from Collaborative Learning Environment. In *Proceedings of the 10th International Conference on Educational Data Mining*, 104–111.
[5] M.T.H. Chi, P.J. Feltovich, and R. Glaser. 1981. Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science* 5, 2: 121–152.
[6] S.A. Crossley, K Kyle, J.L. Davenport, and D.S. McNamara. 2016. Automatic assessment of constructed response data. In *EDM*, 336–340.
[7] P.F. Ericsson and R. Haswell (eds.). 2006. *Machine scoring of student essays: Truth and consequences*. Utah State University Press, Logan, UT.
[8] B. G. Glaser and A. L Struss. 1999. *The discovery of grounded theory: Strategies for qualitative research*. Aldine de Gruyter, New York.
[9] C Glesne. 1999. *Becoming Qualitative Researchers: An Introduction*. Longman, New York.
[10] A.C. Graesser, D.S. McNamara, M.M. Louwerse, and Z Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 36, 2: 193–202.
[11] I Howley and C.P. Rose. 2016. Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics* 3, 3: 239–262.
[12] J.W.Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin, Austin, TX.
[13] D.W. Shaffer. 2006. *How Computer Games Help Children Learn*. Palgrave Macmillan, New York.
[14] D.W. Shaffer. 2017. *Quantitative Ethnography*. Cathcart Press, Madison, WI.
[15] D.W. Shaffer, W. Collier, and A.R. Ruis. 2016. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics* 3, 3, 9-45.
[16] M. Warschauer and D. Grimes. 2008. Automated writing assessments in the classroom. *Pedagogies: An International Journal* 3, 1: 22–36.
[17] A.F. Wise and D.W. Shaffer. 2015. Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics* 2, 2: 5–13.