Finding Common Ground: A Method for Measuring Recent Temporal Context in Analyses of Complex, Collaborative Thinking

Andrew R. Ruis, University of Wisconsin–Madison, arruis@wisc.edu
Amanda L. Siebert-Evenstone, University of Wisconsin–Madison, alevenstone@wisc.edu
Rebecca Pozen, University of Wisconsin–Madison, rnpozen@gmail.com
Brendan R. Eagan, University of Wisconsin–Madison, beagan@wisc.edu
David Williamson Shaffer, University of Wisconsin–Madison & Aalborg University, david.shaffer@wisc.edu

Abstract: Complex, collaborative thinking is often conceptualized as a process of developing cognitive connections among the contributions of different participants. A central problem in modeling collaboration in this way is thus determining, for any contribution to a discussion, the appropriate context for modeling the connections being made—that is, for determining the appropriate recent temporal context. Recent temporal context is typically defined using a moving window of fixed length. However, that length is dependent on the setting, and there are no existing methods for reliably determining an appropriate window length. This paper presents an empirical method for measuring recent temporal context, and thus for defining an appropriate window length to be used in analyses of complex, collaborative thinking. Importantly, the method we describe minimizes the need for human annotation while providing both qualitative and quantitative warrants for choosing a particular window length.

Introduction

In the learning sciences, complex thinking is often conceptualized as a process of developing cognitive connections among concepts (DiSessa, 1988; Linn, Eylon, & Davis, 2004; Shaffer, 2012). In computer-supported collaborative learning (CSCL) contexts, individuals make such connections not only within their own contributions but also to the contributions of their collaborators (Garrison, Anderson, & Archer, 2001; Shaffer, 2017). A central problem in modeling complex, collaborative thinking in terms of cognitive connections is thus determining, for any contribution to a discussion, the appropriate temporal context for modeling the connections being made. Prior work in the learning sciences has approached this problem using *moving windows* (Dyke, Rohit Kumar, Hua, & Rosé, 2012; Siebert-Evenstone et al., 2016), where each turn of talk is associated with some prior segment of the discussion that forms its *recent temporal context* (Suthers & Desiato, 2012).

In such models, analysis of a given turn of talk accounts for both its own content and the content of its associated window. Because there are many variables that may affect the extent of the recent temporal context, including domain, topical complexity, age of the participants, and communication medium, it is important to identify an appropriate window length for each setting. However, there are no reliable methods for measuring the extent of recent temporal context, and studies have not been conducted that show the effects of window length on the features or interpretation of window-based learning analytic models. In this study, we present an empirical method for measuring the extent of recent temporal context. We then evaluate the method by analyzing conceptual connectivity in the same dataset using different window lengths to explore the effects of window length on the resulting models. The results suggest that an appropriate moving window length can be empirically determined with minimal effort, and that while window length can significantly affect model features and interpretation, the empirical method we describe produces relatively robust models of complex thinking.

Background

Collaborative learning takes place both in face-to-face interactions and remotely, mediated by various communications technologies. In addition, computer simulations, intelligent tutoring systems, educational games, and other CSCL learning technologies create settings in which students work with one another, with educators or mentors, and with pedagogical agents to frame, investigate, and solve complex problems. Such settings foster the development of *communities of inquiry* (Garrison & Akyol, 2013; Seixas, 1993): groups of participants who interact with one another in a problem-solving context to facilitate critical discourse and establish shared meaning and understanding. A key goal of such collaborative interactions is to help participants make explicit connections among the contributions of different team members, which enables communities of inquiry to organize knowledge, identify and address misconceptions, facilitate reflection, and ultimately construct mutual understanding (Garrison et al., 2001).

Modeling complex, collaborative thinking, then, requires assessing not only the connections that individuals make among their own contributions but also the connections they make to the contributions of the other participants. This in turn requires identifying the appropriate relational context for any given turn of talk in a discussion (Arvaja, Salovaara, Häkkinen, & Järvelä, 2007). Researchers often establish relational context by looking at each turn of talk and identifying its *referents*: the prior turn or turns of talk that provide the information needed to understand the meaning of a given utterance (Shaffer, 2017).

Illustrating the problem

Consider the following excerpt, in which a student project team working on a biomedical engineering design project is discussing with a mentor (Maria) how they determined which design prototypes to test (emphasis added).

1	Maria	What role did the requests of the **internal consultants** play in deciding which prototypes to test?
2	Jill	thats it though. then we have to pick a final one and make a poster presentation
3	Jill	**Internal consultants** made it clear **reactivity** is important, but **cost** triumphs
4	Lesenia	the **requests** gave a focus on what **factors** to pay attention to
5	Maria	How did you design your device to address **patient** needs?
6	William	We tried to make them **cheap** and **reliable** so they could last a long time and save the **patients** money
7	Brad	Since we are trying to **make a bunch of people happy**, we wanted a wide **variety of results** to see which one **makes the most people happy**.

Maria asks the students a question in Line 1, which Jill and Lesenia answer in Lines 3 and 4. Maria then asks a second question in Line 5, to which William responds in Line 6. In Line 7, Brad's response references "people" twice, and we can identify who they are only by looking to previous utterances—that is, by identifying the referent(s) for Line 7. Yet how far back do we need to look? If the context for his utterance consists only of Lines 5 and 6, we may determine that Brad is referring to patients, the users of the device the team is designing. However, if all the lines are included as context, it is clear that by "people" Brad means both the patients *and* the internal consultants. Interpretation of Brad's utterance—and by extension, interpretation of the connections that Brad is contributing to the team's mutual understanding—depends on how we define his contribution's recent temporal context. That is, there is some window (in this case, seven turns of talk) that contains all of the context (i.e., all of the referents) needed to understand a given contribution to a discussion.

Ideally, the size of a given window would be determined by the content of the discussion, as the context needed for interpreting any given turn of talk will vary. In the excerpt above, understanding Brad's contribution requires a window of seven turns of talk, while understanding William's contribution in Line 6 requires only two: he is answering Maria's question in Line 5, while Brad is addressing both of Maria's questions (Lines 1 and 5) and building on the contributions of Jill, Lesenia, and William. Note that if a larger window were used to analyze William's contribution, he would be credited for making connections to the internal consultants and their concerns, which he does not seem to be doing. In other words, a larger window may produce more false positive connections. However, while humans can identify relational context with relative ease, manual annotation of discourse is not a scalable approach. As large volumes of rich process data are increasingly produced by a variety of CSCL environments, rapid and reliable procedures for measuring recent temporal context are necessary for analyses of complex, collaborative thinking at scale.

Existing approaches

Research on natural language processing suggests that identifying changes in topic or turn-by-turn references to prior utterances cannot be reliably automated (Rosé et al., 2008). Even if automated processes were feasible, the excerpt above illustrates a key problem with segmentation by changes in topic. In Line 5, Maria *is* changing the topic—from the needs of the internal consultants to the needs of the patients—but Brad (Line 7) uses that as an opportunity to make a point about how, as engineers, the project team needs to consider the concerns of multiple different stakeholders when designing products. That is, he is making a connection between the two topics (Lines 1–4 and Lines 5–6). Moreover, even if a given utterance does not make explicit reference to any of the immediately preceding turns of talk, those utterances nonetheless form the *common ground* for that part of the conversation and are likely to influence what participants say and do (Clark, 1996).

Thus for both practical and theoretical reasons, researchers typically use a moving window of fixed length (e.g., a particular duration of time or number of turns of talk) to analyze collaborative connection-making in recent temporal context. Instead of creating summary values for all utterances in a conversation, a moving window analysis computes a value for each utterance, based on the content of that utterance and the content of preceding utterances that are contained within the window. This type of analysis has been used to explore intragroup interactions (Dyke et al., 2012) and to study connection-making in collaborative discourse (Csanadi, Eagan, Shaffer, Kollar, & Fischer, 2017; Quardokus Fisher, Hirshfield, Siebert-Evenstone, Arastoopour, & Koretsky, 2016; Siebert-Evenstone et al., 2016; Sullivan et al., 2018; Suthers & Desiato, 2012).

One technique that uses moving windows to model connections in recent temporal context is *epistemic network analysis* (ENA) (Shaffer, 2018; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017). ENA takes interaction data coded for elements of complex thinking and constructs weighted network models that represent the structure of connections made among those elements by each individual in the dataset. For each line in a conversation, ENA computes a network of connections within the recent temporal context as defined by the window length; that is, ENA identifies the co-occurrences of codes between a given utterance and the previous utterances in the conversation that fall within the window. Proceeding line-by-line through the data, ENA accumulates these networks for each individual to model the unique connections contributed by that person to the conversation. Critically, the length of the moving window—in ENA and in other window-based techniques for modeling conceptual connectivity—defines which *co-occurrences of codes* are hypothesized to represent *meaningful cognitive connections*.

The challenge

A key challenge for window-based models of complex, collaborative thinking is thus to determine a fixed window length that is sufficiently long to capture the recent temporal context for most utterances but not so long as to overrepresent connections that are not meaningful. Because the length of the window determines which connections are included in the model, interpretation of the model may be significantly affected by the choice of window length (Gleicher, 2016). Researchers typically make such determinations on a case-by-case basis, usually without detailing the method used, and an extensive literature search revealed no scalable, objective techniques for measuring recent temporal context. In what follows, we present a novel method for determining window length that minimizes the need for human annotation and provides both qualitative and quantitative warrants for using a certain window length to analyze collaborative connection-making in the specified setting.

Methods

Setting and data

In this study, we analyzed the collaborative interactions of students in the engineering virtual internship *Nephrotex* (Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Chesler, Arastoopour, D'Angelo, Bagley, & Shaffer, 2013; Chesler et al., 2015). In *Nephrotex*, students work in project teams to design an ultrafiltration membrane for a hemodialysis system. The virtual internship is divided into a series of activities that simulate steps in the design process, including reviewing research reports; designing prototypes; discussing design choices with teammates and an engineering advisor; and addressing the needs of internal consultants and external clients. Students interact with their teams and with their engineering advisor through an online instant message program (chat), and the system automatically records all chat conversations for subsequent analysis. *Nephrotex* takes approximately 15 hours to complete.

Chat conversations (N = 54,896 chats) were collected from 20 runs of *Nephrotex* at five institutions in the United States. Participants (N = 652) were first- and second-year college students using *Nephrotex* as part of an engineering course. Students were randomly assigned to project teams of 4–5 members.

Window length annotation

To measure recent temporal context in this setting, we randomly selected 200 utterances from the 54,896 chats in the *Nephrotex* dataset. For each chat, two independent raters identified all previous referents in the conversation. These annotations indicate, for each utterance, the window containing that utterance's recent temporal context, where window length is the number of chats from the referring utterance to the earliest referent, inclusive. In the excerpt above, for example, the window length for Brad's utterance is seven lines: the referring chat (Line 7); the furthest referent (Maria's chat in Line 1); and the intervening five chats.

To calculate agreement between the two independent raters, we computed Cohen's κ (kappa) for each window length. Kappa was calculated for each window size, x, by assigning a "1" to any utterance that a given rater determined to have window length x and a "0" to all other utterances. Kappa thus indicated the extent to which the two raters agreed in their assessments of each utterance's recent temporal context. To determine whether kappa scores for the sample (200 chats) could be generalized to the population from which they were drawn (> 50,000 chats), we computed Shaffer's ρ (rho) to estimate the expected Type I error rate of kappa given the sample size (Eagan, Rogers, Pozen, Marquart, & Shaffer, 2016; Shaffer, 2017).

Effects of window length on models of connectivity

We tested our technique for empirically determining window length by analyzing a portion of the *Nephrotex* dataset using ENA (for a detailed description of ENA methodology, see Shaffer, Collier, & Ruis, 2016). Specifically, we used ENA to model data from two implementations of *Nephrotex* (48 students; 5,757 chats) at window length x for each $x \in \{1, 2, \ldots, 13\}$. At each window length, we compared the networks of (a) students using an engineering virtual internship for the first time (novices; $n = 24$), and (b) students using *Nephrotex* after using a different engineering virtual internship (relative experts; $n = 24$).

The data were coded using an automated coding algorithm for five elements of engineering design: (1) *performance parameters*: the functional attributes of a design; (2) *design decisions*: the process of making design choices, including prioritization and tradeoffs; (3) *client requests*: considering the concerns and needs of stakeholders; (4) *data*: considering technical or numeric information; and (5) *collaboration*: facilitating inclusivity and teamwork in the design process. Inter-rater reliability was computed for each code separately, and all codes had a kappa value greater than 0.75; each kappa value was statistically significant at $\alpha < 0.05$ and $\kappa > 0.65$: $N = 200$, $\kappa > 0.75$, $\rho(0.65) < 0.05$.

Results

Measurement of recent temporal context

Of the 200 chats examined, 49 (24.5%) made no reference to prior chats, and 51 (25.5%) referenced only the previous chat. Thus, a window length of two chats would capture the relevant connections for 50% of the utterances in the sample. However, as Table 1 shows, it is not until a window length of seven that the relevant connections were captured for more than 95% of the utterances. Subsequent increases in window length resulted in only very small improvements, and no utterance required a window length of more than 18 chats.

Table 1: Number and proportion of utterances with complete recent temporal context at each window length

Window Length	Number (and Percentage) of Utterances with Complete RTC	Increase in Percentage of Utterances with Complete RTC over Previous Window Length	Cohen's κ	Shaffer's ρ (for $\kappa > 0.65$)
1	49 (24.5%)		0.96	**< 0.01****
2	100 (50.0%)	+25.5%	0.95	**< 0.01****
3	131 (65.5%)	+15.5%	0.96	**< 0.01****
4	158 (79.0%)	+13.5%	0.97	**< 0.01****
5	170 (85.0%)	+6.0%	0.88	**< 0.01****
6	182 (91.0%)	+6.0%	0.84	**< 0.01****
7	192 (96.0%)	+5.0%	0.94	**< 0.01****
8	195 (97.5%)	+1.5%	0.91	**0.03***
9	195 (97.5%)	+0.0%	0.91	**0.02***
10	197 (98.5%)	+1.0%	0.85	0.11
11	197 (98.5%)	+0.0%	0.66	0.39
12	198 (99.0%)	+0.5%	0.80	0.22
13	198 (99.0%)	+0.0%	0.80	0.22

To calculate the level of agreement between the two raters and to determine whether the findings for the sample (200 chats) are generalizable to the population from which the chats were sampled (> 50,000 chats), we computed kappa and rho for each window length (see Table 1). For all window lengths up to nine, agreement between the two raters was statistically significant for $\kappa > 0.65$ (N = 200, $\kappa \geq 0.84$, $\rho(0.65) < 0.05$), which indicates that the level of agreement between the two raters would have been $\kappa > 0.65$ for those window lengths had they evaluated the entire dataset. While the kappa scores obtained for window lengths greater than nine were not statistically significant at a Type I error rate of $\alpha < 0.05$ (i.e., $\rho(0.65) > 0.05$), this is due to the extremely low number of utterances that refer to chats more than 8 turns earlier. Because the goal of this method is to identify the shortest window that captures the full recent temporal context for most utterances, a window length of seven appears to be optimal for this dataset.

Effects of window length on model features and interpretation

Figure 1 shows the ENA models for moving window lengths one through four (MW1–4). The plotted points (top row), each of which represents the network of one student, show increasing discrimination between novices (red) and relative experts (blue) with increasing window length. The difference graphs (bottom row) compute the difference in connection strengths between the mean networks of the novices and relative experts, showing which connections were stronger in which group. Note that the network graph for MW1, which includes only connections made within individual utterances, is very different from the network graph for MW2, which includes both connections within a given utterance and connections to the immediately preceding utterance. With increasing window length, the network graphs become more stable, as does the interpretation of the model. As the difference graphs for MW3 and MW4 indicate, the novice students made stronger connections among *data*, *performance*, and *design decisions*, while the relative expert students made stronger connections between *client requests* and both *performance* and *design decisions*. In other words, the models for those window lengths suggest that novices were more likely to focus on the technical aspects of the design problem, while relative experts were more likely to focus on the needs of stakeholders in the design process.

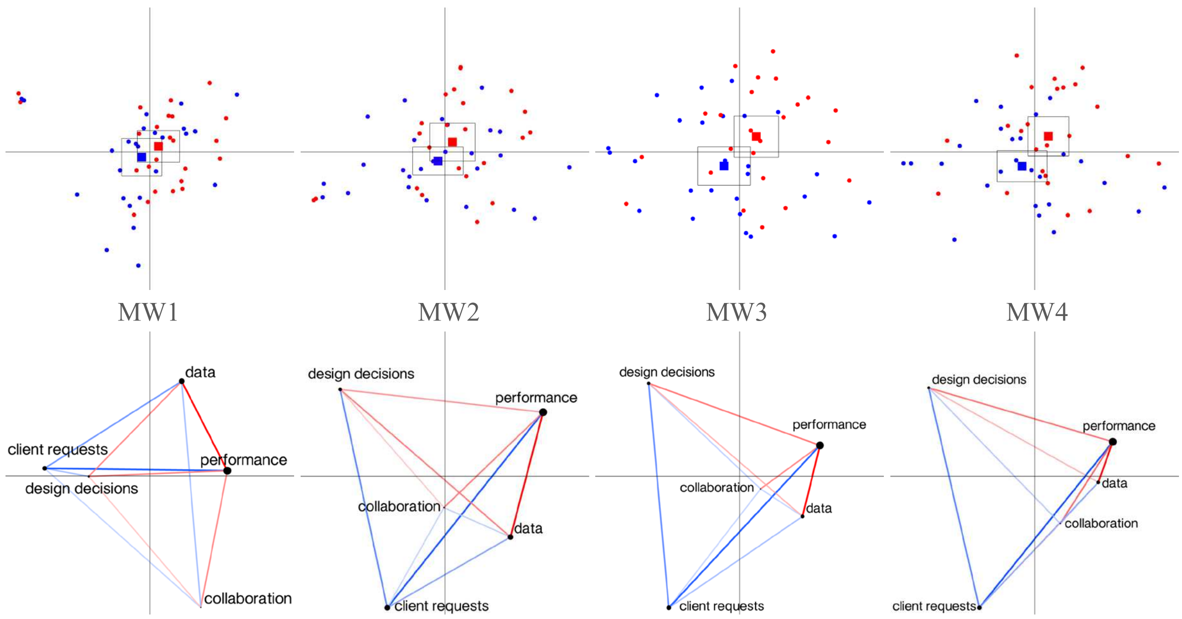

Figure 1. ENA models of the same *Nephrotex* data at four moving window (MW) lengths. The plotted points (top row) show the network locations of the novices (red) and relative experts (blue), with the corresponding means (colored squares) and 95% confidence intervals (boxes). The difference graphs (bottom row) show the difference in connection strength between the means of the two groups.

The summary statistics for the ENA models at all window lengths tested (MW1–13) are shown in Table 2. With the exception of MW5, all ENA models produced using a moving window larger than two indicate a significant difference ($p < 0.05$) between novices and relative experts with a moderate-to-large effect size ($d > 0.60$). However, at a window length of seven, which was hypothesized to be optimal based on our qualitative analysis, the difference between the two groups is significant at $p < 0.01$ for the first time, and the effect size increases almost 10% from the window length of six. The effect size increases further at MW8, but all window lengths from MW9 to MW13 produce summary statistics comparable to MW7.

Figure 2 shows the ENA model for MW7. Note that the both the discrimination between the groups (left) and the structure of connections (right) are similar to the MW4 model shown in Figure 1. While the ability of ENA to discriminate between novices and relative experts is relatively robust to window length, interpretation of the model depends in part on the position of the network nodes in ENA space. The nodes are positioned based on the patterns of connectivity, which in turn are affected by changes in window size. Thus, differences in node positions between models with different window sizes indicate potentially different interpretations of complex, collaborative thinking (1).

Table 2: ENA model discrimination between two study populations at each window length

Window Length	*t*	*p*	Cohen's *d*
1	0.91	0.37	0.26
2	1.58	0.12	0.45
3	2.15	**0.04***	0.62
4	2.46	**0.02***	0.72
5	1.55	0.13	0.45
6	2.44	**0.02***	0.71
7	2.75	**< 0.01****	0.79
8	3.01	**< 0.01****	0.87
9	2.83	**< 0.01****	0.82
10	2.75	**< 0.01****	0.80
11	2.71	**< 0.01****	0.79
12	2.68	**0.01***	0.78
13	2.70	**0.01***	0.79

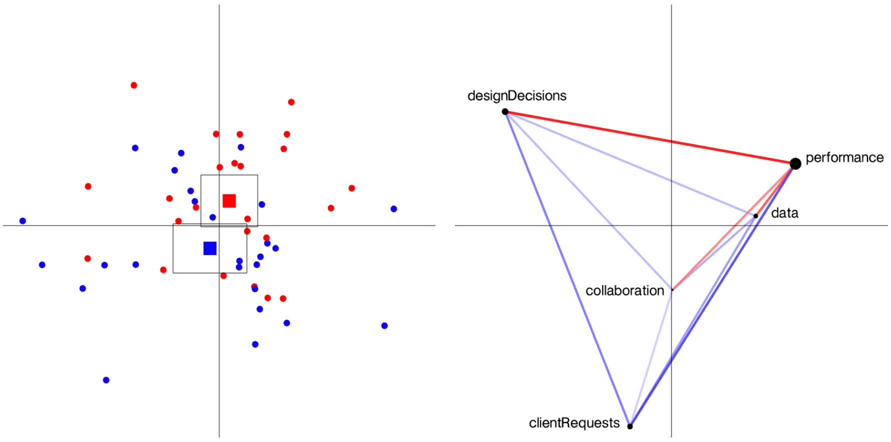

Figure 2. ENA model of the *Nephrotex* data at MW7.

To assess at what window length the network node positions stabilize, we plotted the locations of the nodes at each window length for both the first (*x*) and second (*y*) dimensions of the ENA model (see Figure 3).

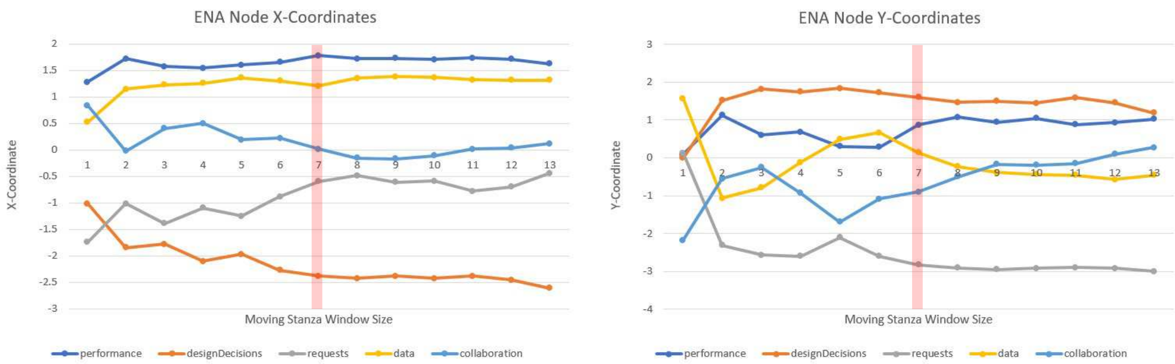

Figure 3. Node positions (*x* and *y* coordinates) in the ENA models at each window length.

The *x*-coordinates of the nodes maintain the same relative order starting at MW2, and the relative spacing stabilizes starting at MW7. After MW7, there are no significant changes in either the relative order or the relative spacing of the nodes. However, the relative order of the *y*-coordinates of the nodes does not begin to stabilize until MW7. D*ata* (yellow) and *collaboration* (light blue) change position relative to one another between MW8 and MW9, but the difference is small and both nodes remain near the origin; thus, the change does not affect interpretation of model. A moving window length of seven is thus where the node positions in ENA space stabilize on both dimensions. Window lengths greater than seven do not add significant information to the model, and window lengths less than seven show different patterns of connectivity, even though models using smaller window lengths do discriminate between the novice and relative expert students.

Discussion

Our goal was to develop a method that (a) provides both qualitative and quantitative warrants for determining the optimal window length for use in moving window analyses of conceptual connectivity in CSCL contexts, while (b) minimizing the number of items requiring human evaluation. To assess this approach, two independent raters analyzed a random sample of 200 student chats (< 0.01% of the 54,896 chats in the dataset). This method identified MW7 as the optimal window length for analyzing these data. We then constructed ENA models of the data that differed only in the choice of window length. This analysis confirmed that a model with a window size of 7 both (a) provides statistical discrimination between groups known to exhibit different patterns of conceptual connectivity based on prior research (Chesler et al., 2015) and (b) provides a stable interpretation of the ENA model. This analysis suggests that statistical discrimination between two groups may be fairly robust to window length once some minimum length is reached (in this case, after MW3 with the exception of MW5), but that model features and interpretation may be more sensitive to window length (the features and interpretation of the ENA models do not fully stabilize until MW7).

As a result, we argue that annotating a subset of data for *furthest referents* makes it possible to analyze recent temporal context and thus determine an appropriate window length to be used in analyses of complex, collaborative thinking. Importantly, this method minimizes the need for human annotation while providing both qualitative and quantitative warrants for choosing a particular window length. While we describe this method by presenting results from one CSCL context (*Nephrotex*) and one learning analytic technique (ENA), we believe that approaches based on annotating a sample data for furthest referents will be compatible with different CSCL settings, different theories of collaborative discourse, and different methods for modeling conceptual connectivity using moving windows. Of course, future research should test our method by repeating this study using other data and other learning analytic models.

While there are many avenues for additional research, this study suggests that hand annotation of a relatively small number of utterances can be used to measure recent temporal context. Critically, this method provides a warrant for making generalizations to the population from which the hand-annotated sample was drawn, making it suitable for analyses of complex, collaborative thinking at scale.

Endnotes

(1) The goodness of fit measures are high (both Spearman's and Pearson's $r > 0.95$) for all 13 models except MW1 ($r \geq 0.85$), which means that the relative positions of the network nodes provide good interpretations of the differences between networks in the model.

References

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501.

Arvaja, M., Salovaara, H., Häkkinen, P., & Järvelä, S. (2007). Combining individual and group-level perspectives for studying collaborative knowledge construction in context. *Learning and Instruction*, *17*(4), 448–459.

Chesler, N. C., Arastoopour, G., D'Angelo, C. M., Bagley, E. A., & Shaffer, D. W. (2013). Design of a professional practice simulator for educating and motivating first-year engineering students. *Advances in Engineering Education*, *3*(3), 1–29.

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, *137*(2), 024701:1-8.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Csanadi, A., Eagan, B., Shaffer, D. W., Kollar, I., & Fischer, F. (2017). Collaborative and individual scientific reasoning of pre-service teachers: New insights through epistemic network analysis (ENA). In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access*

in CSCL: 12th International Conference on Computer-Supported Collaborative Learning (Vol. I, pp. 215–222).
DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 47–70). Hillsdale, NJ: Erlbaum.
Dyke, G., Rohit Kumar, R., Hua, A., & Rose, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In *Proceedings of the 10th International Conference of the Learning Sciences* (pp. 363–370).
Eagan, B. R., Rogers, B., Pozen, R., Marquart, C., & Shaffer, D. W. (2016). rhoR: Rho for inter rater reliability (Version 1.1.0). Retrieved from https://cran.r-project.org/web/packages/rhoR/index.html
Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M. G. Moore (Ed.), *Handbook of distance education* (Vol. 3, pp. 104–120). Routledge.
Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23.
Gleicher, M. (2016). A framework for considering comprehensibility in modeling. *Big Data*, *4*(2), 75–88.
Linn, M. C., Eylon, B.-S., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 29–46). Mahwah, NJ: Lawrence Erlbaum Associates.
Quardokus Fisher, K., Hirshfield, L., Siebert-Evenstone, A. L., Arastoopour, G., & Koretsky, M. (2016). Network analysis of interactions between students and an instructor during design meetings. In *Proceedings of the American Society for Engineering Education* (p. 17035). ASEE.
Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237–271.
Seixas, P. (1993). The community of inquiry as a basis for knowledge and learning: The case of history. *American Educational Research Journal*, *30*(2), 305–324.
Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. D. Squire, & S. A. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 403–431). Cambridge, UK: Cambridge University Press.
Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
Shaffer, D. W. (2018). Epistemic network analysis: Understanding learning by using big data for thick description. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 520–531). New York, NY: Routledge.
Shaffer, D. W., Collier, W., & Ruis, A. R. (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.
Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.
Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. In C.-K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. I, pp. 631–638).
Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., … Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, *163*(4), 938–943.
Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *45th Hawaii International Conference on System Science* (pp. 3368–3377). IEEE.

Acknowledgments

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), 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.