Context Aware CSCL: Moving Toward Contextualized Analysis

Sean P. Goggins, Drexel University, outdoors@acm.org
James Laffey, University of Missouri – Columbia, laffeyj@missouri.edu
Chris Amelung, Washington University in St. Louis, amelungc@gmail.com

Abstract: Groups change over time. CSCL groups, unlike face to face groups, leave evidence of their interactions behind in the form of system logs. Most CSCL log analysis is opportunistic, relying on electronic traces which do not provide information about use context, user content reading behavior or insights about relations among users. To enable context rich analysis of student interactions, we developed a context aware notification system (CANS). In this paper we describe how such logs are processed and analyzed to support the development of multi-mode social networks. In prior studies we reported on analysis of these networks from CANS logs and context enriched logging systems focused on the small group unit of analysis. The purpose of this paper is to increase understanding of the methods we use within the CSCL community. We use CANS logs to create awareness of the social experience of online learning and emergent group formation.

Introduction

In this concept paper we assert that theories of interaction in CSCL must incorporate a more explicit approach to the design of logging systems. Suthers (2010) describes contingency graphs as analytical mechanisms for the study of uptake in CSCL environments in a manner that does not place time at the center of analysis, but instead focuses on conditions that precede and follow important learning acts. Reimann (2009) focuses on the central importance of time, highlighting a gap in our understanding of how time is understood in long running, asynchronous interaction. Stahl (2006; Stahl, 2009b; Stahl, 2009c) focuses on the small group unit of analysis as he calls for a study of the science of group interaction (Stahl, 2009a). Our contribution exists at the intersection of these interests and efforts. We integrate Stahl's focus on the small group unit of analysis, Reimann's attention to time, Suther's analytic view of data and four years of experience using rich, context aware logs to study asynchronous, completely online CSCL environments (cited throughout). With the rest of this paper we review the literature in CSCL focused on advancing the analytical tools available in the field, describe our socio-technical framework (CANS) for more closely connecting interactions to the lived experience of users, and discuss the implications for CSCL theories.

Literature Review

Suthers (2006) called for recognition that ongoing work in CSCL requires integration of design based, phenomenological, and experimental methods to build a complete picture of the intersubjectivity of computer supported collaborative learning. Each methodological tradition is, by itself, too narrow to support the ongoing examination of the socio-technical CSCL experience that we now recognize to be munificent in its variation, even in the same population using the same set of tools over time. Specifically, Suthers (2006) points out that experimental approaches embody an artificiality that constrains our view of how learning events actually happen in the world, design based approaches take the form of iterations and lead to an emergent technomethodology (Dourish & Button, 1996) that practically celebrates munificent variation in uptake and use of CSCL ICT's, and more purely phenomenological studies of CSCL provide descriptions of what occurred. These approaches often fall short in their attempts to provide useful guidance for the development of interventions in the future.

Reimann (2009) takes another perspective on the range of methods utilized in CSCL by presenting a contrast between coding and counting CSCL events, which he calls variable focused analysis, and sequential analysis of events in CSCL settings. This work raises questions about the applicability of analysis of variables in the complex, real world settings that constitute most CSCL environments and the intersubjective learning processes they support. The observed munificent variations of CSCL experience are, to Reimann, an irreconcilable set of multivariate confounds.

As a solution this to core challenge in CSCL research, Reimann (2009) posits that analysis of events and event streams will provide a more authentic view of the intersubjective nature of learning in CSCL environments. While the conceptual insight about the importance of event logs is one we agree with, the conclusion that these events are likely to be digested into semantically meaningful process models and process model instances presumes a linearity and consistency of interaction in CSCL environments for which there is no empirical evidence. Reimann, Frerejean & Thompson (2009) test the idea of applying a process model to event data, concluding that the decision processes in fact take a different path each time. Goggins, Laffey & Tsai (2007), Goggins, Laffey & Galyen (2010a) & Goggins et al (2010b) go further in demonstrating the munificent variations of interaction revealed by event logs. These studies show that even in the most controlled CSCL environments, no two groups follow the same processes. Specifically, those CSCL participants who experience the same CSCL curriculum in the same CSCL, socio-technical environment with the same instructor produce interaction logs with highly variable activity levels, activity sequences and following group-specific processes. This holds for groups who are able to see the activity of other groups and would therefore be at least partially susceptible to social comparison influences (Festinger, 1954) and those where groups are not able to see each others activity (Goggins et al., 2011, International Journal of Computer Supported Cooperative Work, Under Reveiew). Put simply, the notion of nascent, identifiable process models emerging across instances of a CSCL environment is not supported by any data we are aware of.

An alternative to process modeling is Suthers et al's (2007) concept of an eclectic model for examining the interactivity of participants in a CSCL environment. Unlike process models, the eclectic model works to incorporate multiple perspectives from different data. The events themselves are not any more richly constructed than Reimann's (2009), but they are integrated with other data to tell the story of CSCL experience from multiple methodological perspectives. Suthers et al (Suthers et al., 2007) explicate the construct of a dependency graph, which they use as an analytical boundary object for integrating event log data with data from other data types and research methods commonly used in CSCL research. Like Reimann, Suthers et al arrive at a method of examining behavior in logs that relies on establishing a more defined, non-dynamic picture of interaction than what is born out by much experience in the socio-technical systems that constitute CSCL in the wild. The important contribution of the dependency graph in CSCL research is that it facilitates consistent integration of data from the diverse set of research traditions used to examine CSCL.

Eclectic modeling and dependency graphs used as boundary objects raise the question of whether or not existing systems for logging CSCL activity provide sufficient contextual data to support an automated approximation of a dependency graph. Is it possible more context data could be captured in logs than is commonly the case in CSCL today? If we did this, could the automated generation of dependency graphs that are more easily integrated with other forms of CSCL data be realized? The interactive logs we analyze provide the bidirectional view of interaction through technology in a learning environment that is not available for the analysis methods proposed by Reimann (Reimann et al., 2009) and Suthers (Suthers et al., 2007). The logs produced by the CANS system (our unique logging system, described under study context and methods), in fact, provide a wider swath of interaction data for analysis, and as a consequence greater potential for automated analysis of interactivity in computer mediated learning settings. For example, our preliminary analysis of interaction logs over three years of CANS data shows that these sorts of passive, invisible interactions between members are more common than post behaviors by a ratio of 15:1 (6.6% of activity in ~670,000 events is active, posting or creating activity). To understand the uptake of ideas in these environments, CANS logs provide a wide foundation of interactivity records that we will use for the development of insight from events, event vocabularies and event grammars (event vocabularies and grammars are discussed in more depth later in this section).

Our work seeks to elaborate on Suthers et al's (2007) notion of the dependency graph by incorporating a more complete logging infrastructure, and analysis that recognizes the evolving social structures and patterns that can be made visible through these logs. Most significantly, the logs we use capture both passive (reading) and active (posting) behavior of participants. Both Reimann & Suthers et al's approaches rely on the analysis of interactions and interaction logs that only contain a record of the proactive posting behavior of participants. This proactive posting behavior, in response to others and sometimes starting from scratch (as in a new forum) describes the observable, creative acts of participants. Prior research in online awareness (Carroll, Neale, Isenhour, Rosson, & McCrickard, 2003; Carroll, Rosson, Convertino, & Ganoe, 2006; Amelung, 2007; Laffey, Amelung, & Goggins, 2009) across multiple contexts suggests that knowledge of the social presence (Erickson & Kellogg, 2000) of others influences interaction. In CSCL environments, knowledge of who is reading the contributions of which other participants permits the researcher to observe the full intersubjective nature of interactions, and tease out the vocabularies and grammars of interaction which correspond with different levels of performance.

Our work views this less active behavior as a significant but typically invisible indicator of uptake in CSCL environments. As an example, imagine an individual in an online graduate student course who is contributing to a discussion in her small group. The participation in this course occurs in course wide discussion boards and through discussion boards restricted to each small group. Our subject is participating in a design activity with her small group while at the same time contributing to a question posed by the instructor of the course, asking students to list a design researcher whose work is exemplary of the type of work she would like to do. If we know that the student is reading the posts of other students in the larger context consistently before she makes contributions to her small group, we have evidence that the meaning making (uptake) is influenced not only by her direct response to her group, but also by her sequential review of other material in the course. If we extend our view of the uptake of information to include other interactions of the member with information resources or colleagues outside the course, it quickly becomes apparent that our view of uptake is materially influenced by the width of data we capture.

Our work adapting Suthers et al's (2007) notion of uptake emerges as a two-step interpretive dance. The first step is analyzing sets of log data that permit a researcher to witness emergent social structures, combined with the full indexicality of socio-technical interactions – active and passive. This will potentially change both our view of the nature of mixed methods research in CSCL, but also expose the potential richness of member event grammars that may be constructed from these logs. As we noted earlier, passive actions outnumber active actions in the systems we have studied by a ratio of 15:1. Our logs provide 15x more data for constructing event grammars. As a second step outside the scope of this paper, but important to our long term research agenda, capturing log data from multiple socio-technical contexts to reflect the actions of a group of CSCL users will widen our view of indexicality and possibly introduce a new set of CSCL patterns.

For electronic trace data, the small group unit of analysis truly becomes, as Stahl (2006) states, "Where the action is" in CSCL. Traces measure interactions between a user and a system; CANS captures more context data than any other system and from that data we can reconstruct the multi-modal social network experienced by students and instructors.

Which Log Structures are Analytically Useful?

Conversation analysis from the ethnomethodological tradition and social network analysis from structural sociology contribute theoretical and methodological perspectives to our work defining log analysis and log design in CSCL. The context data and completeness of the interaction record that is revealed by CANS log data is also foundational to our ability to proceed with a study that explores the utility of log and sequence structures for CSCL research. Our goal here is to discover new ways of analyzing electronic traces by identifying semantically meaningful grammars from event trace data in CSCL environments. CANS is the first system we are aware of which has collected at least five years of rich trace data from completely online learning groups.

Research methods like conversation analysis have examined the microstructure and indexicality of conversation to explicate meaning. The events we record and grammars we identify are derived from socio-technical interactions recorded in CANS event logs from an online course management system. While conversation analysis has a long tradition of examining micro-sequences of conversation to explicate meaning and understand the perspective of conversational participants, our method is new. In conversation analysis, the sequential order of utterances and non-verbal communication are viewed together with a particular expression, word or phrase to discern meaning. In CSCL, the Virtual Math Teams research led by Gerry Stahl (Stahl, 2006; Cakir, 2007; Stahl, 2009c; Stahl, Ou, Cakir, Weimar, & Goggins, 2010) demonstrate the applicability of methods derived from conversation analysis and ethnomethodology to the analysis of socio-technical conversations around objects in a synchronous CSCL space.

We show that insights about structural change and interesting events emerge from our log analysis, and suggest that CSCL researchers who use conversation analysis will benefit from first applying our methods as a filter for selection of specific events for coding. Readers should note that we think that obtaining the richness of understanding that emerges from ethnomethodolically informed methods, like those used on the VMT project through automated means, is an unrealistic goal. Without the benefit of close, interpretive examination of discourse there is certain to be loss in meaning, which is a natural outcome of choosing a vocabulary of events and event grammars, which are required steps in the production and analysis of electronic trace data.

To begin to accomplish an approximation of the understanding derived through conversation analysis from automated analysis of logs, we need to identify meaningful structures and sequences (grammars) from the available logged events (vocabularies). The construction of semantically meaningful events begins with a vocabulary of logged event types, and a search for social structural patterns (who relates to who), the trajectory of evolution in social structure and a complementary analysis of interaction sequences across the socio-technical context, which includes the layers of the system context. Events, however they are defined, are initially captured at some imperfect, but consistent granularity that works to mute subtle gestures, references and turn-taking activities that constitute the actual, qualitative experience of conversation in the electronically mediated or real world.

Event grammars built up from event vocabularies do not hold enough information to permit us to approximate the results of conversation analysis research methods through automated log analysis. We must also connect performance within groups to the structural patterns and sequences of interaction made visible through our bidirectional usage logs in order to connect events to our understanding of uptake in CSCL.

The analyzed events then become inputs to a multi-modal social network, which can be analyzed using context enhanced forms of social network analysis, which we describe in Goggins, Laffey & Gallagher (2011)

Trace Data Resolution Levels

Our insight, derived from years of analysis and active participation in the field of CSCL suggests that how logs are structured at runtime, and for analysis, can represent the lived interactions of users through technology to greater and lesser degrees. In table one, we propose four data resolution levels for CSCL researchers to consider. The lowest level simply includes a user id, environment code, context id, session id, url, event type (read, post, etc), event object and a timestamp. Events are captured serially from the CANS system, and modified to support this analysis in each subsequent level. The expanding structure of capture and analysis described in table one is somewhat abstract, but reflects a clear description of how each progressive step provides the CSCL researcher with a more fine grained view of the experienced interactions of users. Access to the corresponding, generalized toolset will be made available as part of a conference presentation.

Table 1: Data resolution levels for CANS data analysis.

Data Resolution Level	**Data Resolution Level Description**
(1) Raw CANS Data	One event per row.
(2) Bi-Directional CANS Data	For example, if I read a discussion board topic that you created, then a connection is drawn between you and I. In addition to the data in "Raw CANS Data", this data set contains: 1. The distance in minutes between an event and the object (usually a discussion board) that the event is in response to. 2. An identifier for the object creator; this creates the social link.
(3) Exploded Bi-Directional CANS Data	Exploded Bi-Directional CANS data, as explained in Goggins, Laffey, Amelung & Gallager (2010b) and referenced in Goggins, Galyen and Laffey (2010a) recognizes the social form of online discussion. This includes recognition that when an individual participates in a discussion board or other interaction in a CSCL or other collaboration system, they frequently view more than one, specific post. This varies by environment. Effectively, one row is created for each artifact that is visible on a page when the page is viewed. In the systems we study, this is discerned from the timestamp, url and event object (artifact or discussion).
(4) Weighted Extraction of Exploded Bi-Directional CANS Data	The calculated time distance at level three can be transformed into a meaningful weighting factor for the analyst, depending on how other data gathered suggests weights should be calculated. This provides a concrete, automated method to support dependency graph construction with more refined weighting. For example, in many of our studies, interviews and field notes suggest a 3 to 4 day "cliffing" of the interaction weights is appropriate. These weighting "cliffs" depend on the environment studied.

Implications for Theory and Conclusion

The log capture and design techniques that we conceptualized in this paper provide a representative view socio-technical context in CSCL, and advance a broad range of CSCL research agendas. Data captured by most systems is designed for the convenience of system analysts and web based metrics, not for the analysis of social, collaborative behavior. CSCL log data should analytically useful, and it should represent social interactions that are both implicit and explicit. Time matters because more recent interactions are more salient for measuring the social nature of asynchronous interaction. Activity and context matter for connecting analytical dots, as Suthers, Reimann and Stahl all do. The analysis and capture we propose here takes a more full account of activity. For knowledge construction, performance and other questions in CSCL research, the full social context is critical. For our work, and the work of those in CSCL who have turned toward the social, the log analysis we propose here is a necessary and fundamental shift required when log analysis is a central method. We think our approach complements the recent methodological and analytical innovations of others in CSCL. The techniques and tools we have develop and describe here enable CSCL researchers to more fully incorporate social theories of learning (Bandura, 1977) into their work, because our approach makes the social visible.

References

Amelung, C. (2007). Using Social Context and E-Learner Identity as a Framework for an E-Learning Notification System. *International Journal on E-Learning*, *6*(4), 501-517.

Bandura, A. (1977). *Social Learning Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Blincoe, K., Valetto, G., & Goggins, S. (2011). *Leveraging Task Contexts for Managing Developers' Coordination*. Proceedings from ACM International Conference on Software Engineering, Hawaii.

Cakir, M. P. (2007). *How online small groups create and use mathematical objects to do collaborative problem solving*. Proceedings from Proceedings of the 8th international conference on Computer supported collaborative learning.

Carroll, J. M., Rosson, M. B., Convertino, G., & Ganoe, C. H. (2006). Awareness and Teamwork in Computer Supported Collaborations. *Interacting With Computers*, *18*, 21-46.

Carroll, J. M., Neale, D. C., Isenhour, P. L., Rosson, M. B., & McCrickard, D. S. (2003). Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies*, *58*, 605-632.

Dourish, P., & Button, G. (1996). Technomethodology: Paradoxes and Possibilities. *Proceedings of CHI96 Human Factors in Computing Systems,* 13-18.

Erickson, T., & Kellogg, W. A. (2000). Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction*, *7*, 59-83.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*, 117-140.

Foster, J., & Goggins, S. (2011). Infrastructuring New Media for Education. *Journal of the American Society for Information Science*, *under review*.

Goggins, S., Laffey, J., Galyen, K., & Mascaro, C. (2011). Group Awareness in Completely Online Learning Groups: Identity, Structure, Efficacy and Performance. *International Journal of Computer Supported Cooperative Work*, *Under Review*.

Goggins, S., & Erdelez, S. (2010). Collaborative Information Behavior in Completely Online Groups. In J. Foster (Ed.), *Collaborative Information Behavior: User Engagement and Communication Sharing.* Hershey, PA: ISI Global.

Goggins, S., Galyen, K., & Laffey, J. (2010a). *Network Analysis of Trace Data for the Support of Group Work: Activity Patterns in a Completely Online Course*. Proceedings from ACM Group 2010, Sanibel Island, FL.

Goggins, S., Laffey, J., & Gallagher, M. (2011). Completely Online Group Formation and Development: Small Groups as Socio-Technical Systems. *Information Technology & People*, *Accepted*.

Goggins, S., Laffey, J., & Tsai, I.-C. (2007). *Cooperation and Groupness: Community Formation in Small online Collaborative Groups*. Proceedings from Proceedings of the ACM Group Conference 2007, Sanibel Island, FL.

Goggins, S., & Mascaro, C. (2011). Context Matters: ICT's Effects on Physical, Informational and Cultural Distance in a Rural IT Firm. *The Information Society*, *under review*.

Goggins, S. P., Laffey, J., Amelung, C., & Gallagher, M. (2010b). *Social Intelligence In Completely Online Groups*. Proceedings from IEEE International Conference on Social Computing, Minneapolis, MN.

Laffey, J., Amelung, C., & Goggins, S. (2009). A Context Awareness System for Online Learning: Design Based Research. *International Journal on E-Learning*, *8*(3), 313-330.

Laffey, J., Tsai, I.-C., Amelung, C., Hong, R.-Y., Galyen, K., & Goggins, S. P. (2011). The Role of Social Information for Social Ability, Sense of Community and Satisfaction in Online Learning. *Journal of Open and Distance Education Research*, *Accepted*.

Mascaro, C., & Goggins, S. P. (2011). *Brewing Up Citizen Engagement: The Coffee Party on Facebook*. Proceedings from Communities & Technologies, Brisbane, Australia.

Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239-257.

Reimann, P., Frerejean, J., & Thompson, K. (2009). *Using Process Mining to Identify Models of Group Decision Making in Chat Data*. Proceedings from CSCL 2009, Rhodes, Greece.

Stahl, G. (2009a). *For a science of group interaction*. Proceedings from Proceedings of the ACM 2009 international conference on Supporting group work.

Stahl, G., Ou, J. X., Cakir, M. P., Weimar, S., & Goggins, S. (2010). *Multi-User Support for Virtual Geogebra Teams*. Proceedings from GeoGebra NA2010, Ithica,NY.

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge.* Boston, MA: MIT Press.

Stahl, G. (2009b). Meaning Making in VMT. *Studying Virtual Math Teams* Retrieved July 2, 2009, from http://www.ischool.drexel.edu/faculty/gerry/vmt/book/26.pdf.

Stahl, G. (2009c). Toward a Science of Group Cognition. *Studying Virtual Math Teams* Retrieved July 2, 2009, from http://www.ischool.drexel.edu/faculty/gerry/vmt/book/28.pdf.

Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). *A framework for eclectic analysis of collaborative interaction*. Proceedings from Proceedings of the 8th international conference on Computer supported collaborative learning.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315-337.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5-42.