

A Unified Framework for Multi-Level Analysis of Distributed Learning

Daniel Suthers¹, Devan Rosen²

¹ Department of Information and Computer Sciences, University of Hawaii,
1680 East-West Road, POST 309, Honolulu, HI 96822, USA
suthers@hawaii.edu

² Department of Speech, University of Hawaii, 2560 Campus Road,
George 326, Honolulu, HI 96822, USA
rosend@hawaii.edu

Abstract. Learning and knowledge creation is often distributed across multiple media and sites in networked environments. Traces of such activity may be fragmented across multiple logs and may not match analytic needs. As a result, the coherence of distributed interaction and emergent phenomena are analytically cloaked. Understanding distributed learning and knowledge creation requires multi-level analysis of the situated accomplishments of individuals and small groups and of how this local activity gives rise to larger phenomena in a network. We have developed an abstract transcript representation that provides a unified analytic artifact of distributed activity, and an analytic hierarchy that supports multiple levels of analysis. Log files are abstracted to directed graphs that record observed relationships (contingencies) between events, which may be interpreted as evidence of interaction and other influences between actors. Contingency graphs are further abstracted to two-mode directed graphs that record how associations between actors are mediated by digital artifacts and summarize sequential patterns of interaction. Transitive closure of these sociograms yields sociograms, to which existing network analytic techniques may be applied, yielding aggregate results that can then be interpreted by reference to the other levels of analysis. We discuss how the analytic hierarchy bridges between levels of analysis and theory.

Keywords: Socio-technical networks, distributed learning, networked learning, interaction analysis, social network analysis.

1 Introduction

The rapid adoption of information and communication technologies (ICT) in support of “online,” “distributed,” and “networked” learning and knowledge creation activities [1], and their blending with face-to-face venues [14] is well known to the research community to which this paper is addressed. In this paper we use *learning* as shorthand to include any enhancements of individual or collective knowledge or skills, whether or not it occurs in formal educational settings. We include in our scope of interest learning in (for example) online university settings, professional communities, and virtual organizations [2, 4, 8, 29]. We will refer to these collectively as *socio-technical networks* [19]. A related trend is towards open learning

communities. Courses in formal educational settings need no longer isolate participants from others in different courses, but can embed courses in online communities of learners, for example supporting transdisciplinary graduate education [10]. In corporate or other work settings, professional learning communities similarly may cross team contexts rather than being isolated in work teams [42]. The fundamental question of interest in all of these settings is how learning takes place through the interplay between individual and collective agency. All learning activity requires that individuals take actions, but these individual actions are contingent on the actions of others in their socio-technical network contexts, actions that reflexively construct those contexts.

The first analytic challenge addressed by this paper is that learning and knowledge creation activities in these networked environments are often distributed across multiple media and sites. As a result, traces of such activity may be fragmented across multiple logs. For example, the networked learning environments we study offer mixtures of threaded discussion, synchronous chats, wikis, whiteboards, profiles, and resource sharing. Events in these media may be logged in different formats and recorded in databases and text files, disassociating actions that for participants were part of a single unified activity. This disassociation is exacerbated when activity is distributed across multiple virtual sites or spread over time. Also, the granularity at which events are recorded may not match analytic needs, and media-level events may be the wrong ontology for analyses that begin with relationships rather than individual acts. Translation from log file representations to other levels of description may be required to begin the primary analysis. As a result of these various issues, the coherence of distributed interaction and phenomena that emerge from this interaction are analytically cloaked.

Furthermore, understanding distributed learning and knowledge creation requires multi-level analysis of the situated accomplishments of individuals and small groups and of how these local accomplishments give rise to larger phenomena in networks such as the dissemination and transformation of ideas, implicit coordination of the activities of many participants, and the accrual of collective knowledge. Consider the question of how the design of the virtual environment influences emergent phenomena. Everything builds on the existence of multiple successive moments in which an individual is experiencing some presentation of the virtual environment, cares enough to act, and is able to choose an appropriate action. Whether and how this action has implications for network or community level phenomena requires that some trace of the action be given persistent form that other participants might later encounter in their experience of the virtual environment [18]. Appropriate aggregation and availability of such traces can drive dissemination of ideas, align participants, and lead to accrual of collective resources of value. Critically, an empirically grounded understanding of this emergence requires analysis at both fine-grained and aggregate levels. The same can be said for understanding the relationship between small group interactions and larger scale phenomena.

In summary, since interaction is distributed across space, time, and media, and the data comes in a variety of formats, there is no single transcript to inspect and share, and the available data representations may not make interaction and its consequences apparent. To address these concerns (and to support the diverse research in our laboratory), we have developed a framework consisting of an abstract transcript

representation that collects relevant events into a single analytic artifact, and an analytic hierarchy that supports multiple levels of analysis. This paper describes the framework and discusses its potential roles in unifying multiple sources of data and bridging between levels of analysis and theory. We discuss how the framework addresses several specific analytic needs, including: (a) scaling up microanalysis of interaction to large data sets, (b) enabling the translation of event logs into tie data appropriate for social network analysis, and (c) interpreting results at one level in terms of another (e.g., relating social network analytic results back to their interactional settings). Throughout the paper, a simple example drawn from our prior research illustrates many of the features of the framework.

2 Preview

The analytic hierarchy consists of several abstraction layers of analytic representations that we have found to be useful, summarized in Table 1. *Process traces* such as log files are abstracted to *domain models* describing the actors, actions and media objects involved in *event models*, which are collections of temporally tagged events. These event models can be further elaborated by installing directed graphs of empirical relationships between events called *contingencies*. Contingencies can be any observed relationship between events (e.g., two events are by the same actor, involve the same object, are temporally contiguous or proximal, or overlap in content). Contingencies situate participants’ acts in relation to other events—hence the name *contextualized action* model. The analytic utility of contingency graphs is enhanced if focused on those contingencies that may be interpreted as evidence of *uptake*: interaction and other influences between actors. When such interpretations are made, contingency graphs are abstracted into uptake graphs, representing *interaction models*. Interaction can be further abstracted to two-mode directed graphs, called *associograms*, which record how associations between actors are mediated by their

Table 1. The Analytic Hierarchy

<i>Models</i>	<i>Representations</i>
Process Trace	Log files, audio and video recordings, etc.
Domain	Entities and their relationships (types and instances of both)
Event	Sets of events (described in terms of actors, objects, time, etc.)
Contextualized Action	Contingency graphs indicating empirical relationships (contingencies) between events
Interaction	Uptake graphs (each arc corresponds to bundles of contingencies that evidence uptake)
Mediation	Associograms: two-mode directed graphs relating actors to objects
Relationship	Subgraphs of the mediation model consisting of all paths between two actors
Tie	Sociograms representing ties between actors

creation and modification of and access to digital artifacts: hence the name *mediation models*. Associograms also summarize sequential patterns of interaction, making it easier to localize certain patterns. Reduction of associograms by transitive closure into direct ties between actors yields sociograms, representing *tie models*. Existing network analytic techniques may be applied to sociograms. The results of network analysis can then be interpreted by reference to the other levels of analysis. Thus associograms bridge between interaction analysis and network analysis.

The analytic concepts (e.g., contingencies, uptake, mediated associations, and ties) in this paper are not new. Rather, the value of the framework relies on the fact that they are abstractions of concepts commonly applied in existing analytic practice (e.g., adjacency, edits, replies, etc.) as will be detailed later. Thus the framework is offered to coordinate and augment rather than replace existing analytic practices.

The layers are explained in more detail in the following subsections. The process trace, domain, and event models and transformations between them are likely to be familiar to readers: brief sections on these layers are included for completeness and to provide the foundation and examples for describing subsequent layers. Contingency and uptake graphs and associograms are more unique contributions, so are described in some detail here—see also [37] for extensive discussion of motivations for contingency graphs and examples of their use for uptake analysis. The most abstract layer is covered substantially in the social network analysis literature [e.g., 41], so is described here only in relation to how it is derived from the layer below, and what that vertical relationship enables that would not be possible with direct measurement of ties. Throughout this presentation, applications to the study of learning analytics are discussed. The methods described in this paper have been applied in numerous analyses of data from an online learning environment and from laboratory studies of ICT mediated collaboration. At present we are using these techniques in analyses of SRI's Tapped-In teacher professional community [12, 32], a virtual organization that hosts many thousands of education professionals annually in more than 8,000 user-created spaces that include IRC, threaded discussions, shared files and URLs, and other tools to support collaborative work.

3 Process Traces

Any analysis of interaction begins with a process trace, or record of activity left in the environment and accessible to the researcher. Examples include software log data (software application or server logs), audio and video recordings, and textual transcripts. The analytic hierarchy described herein was originally designed to support analysis of both software logs and video recordings, sometimes in conjunction (e.g., we have analyzed application logs and screen capture of the same application [25, 26]). For learning analytic applications and to emphasize the potential for automated analysis, this paper focuses on software logs, and does not touch on issues of video analysis; see [15, 17] for discussion of such issues.

The analytic hierarchy is illustrated throughout this paper by building on a simplified example taken from one of our online learning community applications, disCourse. The disCourse environment provides threaded discussions, wiki pages,

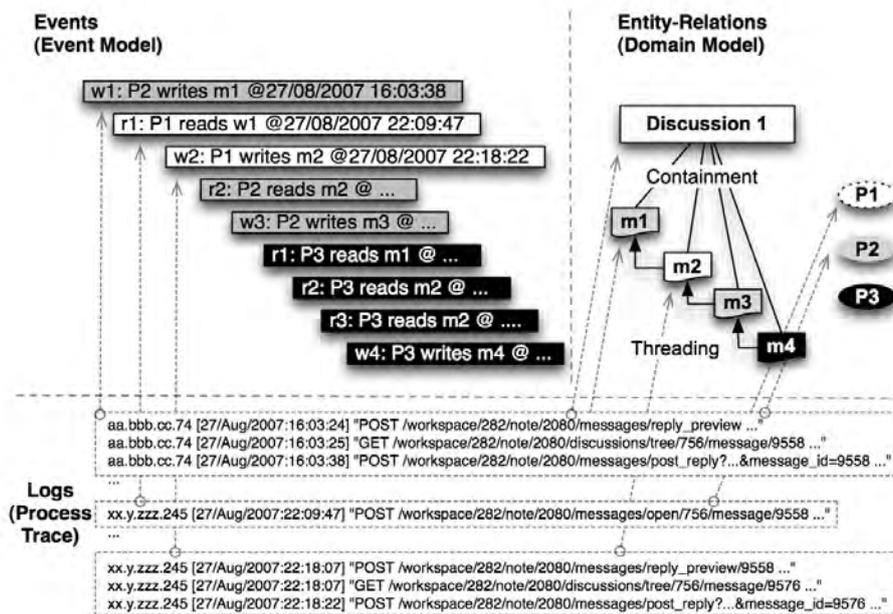


Figure 1. From Process Trace to Domain and Event Models

resource sharing with searchable metadata, and user profiles, organized in a workspace metaphor that collects together tools and resources relevant to a given group, such as a class [36]. The lower portion of Figure 1 shows excerpts (edited for anonymity and simplicity of presentation) of an http server log¹ from disCourse. The example in this paper builds on these logs. See [37] for the full text of the example.

4 Entity-Relations: Domain Model

Prior to or concurrently with the construction of the event model (next section), it is necessary to construct an ontology of the kinds of entities involved in the application domain of interest. Classes of entities and potential structural relationships between them are defined (e.g., actors, discussions, and messages, related by containment, threading and authoring relations). As the trace or log file is processed, new instances of entities and their structural relations are added to the domain model when they are encountered, along with relevant attributes that are expected to be needed for analysis. This is undertaken in conjunction with construction of the event model. For example, the right hand side of Figure 1 illustrates a domain model fragment representing how messages m1, ... m4 are created by participants P1, P2, P3 (shown by shading), related to each other by a threading relation, and contained in a discussion forum. The

¹ disCourse logs events in a database. HTTP server logs of the same events are shown in this example to illustrate the method using log formats familiar to readers.

content of messages are also recorded in the domain model. Temporal information is recorded in the event model, discussed next.

5 Events: Event Model

The process trace is transformed into a set of events that constitute an analysis' first commitments concerning the relevant units for analyzing processes. This transformation involves the Exploratory Sequential Data Analysis (ESDA) operations of *chunking* and *coding* [31]. For example, the first three lines of the log of Figure 1 all are part of the process of posting a message in a system in which each message is previewed before posting. These three traces are chunked together and represented as the single event *w1* in the event model, along with information about the actor (*P2*, indicated by grey), action taken (*w* for writing), object (message *m1*), contents and location (recorded in the domain model), and temporal scope of the action. We call this layer the *event model* because the focus is on individual actions and other events by non-human actants such as software display events. (*Actant* is Latour's [22] term for non-human entities that yet have agency in networks of associations.) The events have not yet been put in relation to each other, other than ordering along a timeline.

Events may be derived from distinct process traces that come from different media, tools or sites, and are recorded in different formats. For example, chat contributions, wiki edits, whiteboard edits, file uploads, etc. can be merged into a single event stream. (To remain faithful to the case example and avoid complicating the figures, this capability is not illustrated in the figures, but it is a simple extension.) A key concern is persistence of identity across tools and sites: some work may be required to ensure that each given actor is represented by the same identifier in the event model, and likewise for the identity of digital objects shared across tools (ideally persistence of identity should be addressed in mash-ups for the learners' sake [20]). Once this has been accomplished, the event and domain models taken together provide an *abstract transcript* of the data that re-assembles in one analytic artifact the diverse events that were for their actors a single activity. If the transcription is complete with respect to the needs of a given analysis, then it is not necessary to retain the original process traces. However, we retain pointers to the original process traces because it may not be possible to identify all needs in advance. We may need to recover other information from the process trace. Also, any transcript includes initial theoretical commitments [11, 28], which may turn out to be faulty, necessitating a return to the original process traces.

A number of analyses can be undertaken on the event and domain models without further analysis. In our research, this is the level at which we answer basic questions about the distribution of activity in the environment: who is participating with whom, in what virtual sites or contexts, and involving what literal content. But to analyze interaction and uncover ties between actors we must relate events to each other.

6 Contingency Graph: Contextualized Action Model

Contingency graphs are an empirically grounded elaboration of the abstract transcript to make analytically relevant relationships between events explicit. We originally called these relationships dependencies, but have renamed them *contingencies* because they capture relationships between events that may be merely contingent or incidental to the situation, rather than being causal or deterministic. The graph simply makes relationships that are latent in the data more explicit, and does not constitute a commitment concerning actors' intentions. Human action can be embedded in its context in many ways, including accidental relationships, or opportunistic leveraging of contextual and historical features as well as necessary antecedents for action [6, 22]. Thus, a contingency graph represents how action is embedded in the context of other events. Examples of contingency types we have used are listed in Table 2 (not intended to be a complete taxonomy). A detailed presentation of the motivations and theory behind contingency graphs and their application to interaction analysis may be found in [37].

Construction of a complete graph of the contingencies between events in a process trace is not practical, as it would result in a graph with a high “signal” to “noise” ratio that is too complex for processing. (Imagine a graph in which each event is linked to every one involving the same actor, or the same object, or that has overlap in lexical content, or occurred nearby in time, and so on.) An analyst chooses those contingencies that are relevant for specific analytic purposes as guided by explicit or implicit theory. Therefore a contingency graph reflects further commitments on the part of the analyst. However, even though a contingency graph is theoretically selective, we always base contingencies on empirically observable relationships between events found in the event and domain models, preferably those relationships that are unambiguous and can be detected automatically. If this standard of evidence is followed, a contingency graph can be treated as an abstract transcript that makes the evidence for interaction or other phenomena of interest manifest.

Table 2. Examples of contingency types

<i>e_k is contingent on e_i when ...</i>	
Media Dependency	<i>e_k</i> operates on a media object or state of that object that was created or modified by <i>e_i</i> (e.g., reply to a message; editing a shared wiki)
Same Actor	<i>e_k</i> and <i>e_i</i> were due to acts of the same actor
Inscriptional Similarity	<i>e_k</i> creates inscriptions with visual attributes similar to those of inscriptions created by <i>e_i</i>
	<i>e_k</i> creates inscriptions with lexical strings identical to those in inscriptions created by <i>e_i</i>
Temporal Proximity	<i>e_k</i> took place soon after <i>e_i</i> , where “soon” depends on the attentional properties of the agents and persistency of the medium (e.g., proximal messages in chat)
Spatial Organization	<i>e_k</i> operates on inscriptions in a spatial context created by <i>e_i</i> (e.g., “grouping” a graphical object by placing it near others)

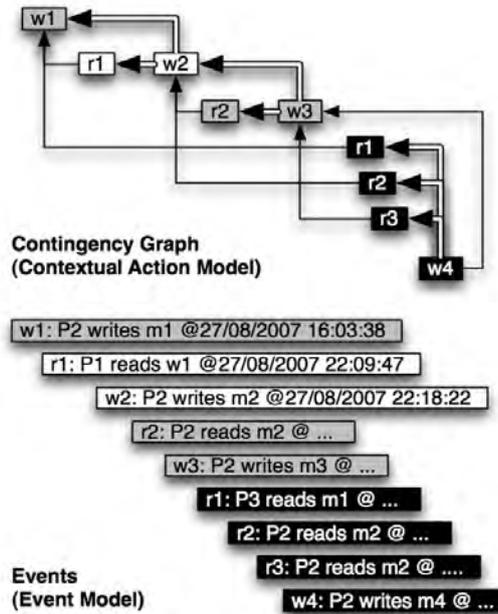


Figure 2. Installing Contingencies

Contingency graphs can be constructed automatically from the layers below it [see, for example, 26]. For example, for each event in which an actor accessed an object we might scan back to find the last event in which the object's contents were modified, and install a media dependency. Contingencies can also be installed from a given event to the most recent prior event involving the actor, to prior events in which the actor accessed a media object with similar inscriptions (e.g., lexical phrases or graphical devices), or to temporally recent events in the same spatial site. A challenge with algorithmic installation of contingencies is limiting their number. Temporal or sequential proximity are useful (and computable) heuristics for selecting relevant contingent events, as they follow the local continuity of human attention and goal directed behavior: what actors do at any given moment is likely to be contingent upon their immediately prior act.

For example, Figure 2 shows the events of Figure 1 with contingencies installed. The single arcs represent media dependencies, and the double arcs represent multiple contingencies, such as temporal proximity combined with same actor and possibly inscriptional similarity. The act of reading a message (r1, r2, etc.) is media-dependent on the act of creating the message (w1, w2, etc.). The act of writing a message (e.g., w2) may be media-dependent on the act of creating the message to which it is a threaded reply (e.g., w1) and is contingent on the messages that the author has recently read (e.g., r1). In this example, the message created by w2 contained a noun-phrase in common with that created by w1.

Once constructed, various kinds of analytic actions are possible on contingency graphs. For example, suppose a particularly productive session was identified in which participants made significant ideational progress. One option is to examine the interaction of the session participants more closely to identify the relationship between group processes and their accomplishments, and how participants appropriated the interactional affordances of the available media for these purposes. We have used the contingency graphs in several studies to support this kind of microanalysis of interaction [24, 25, 26, 38]. Recurring patterns of interaction so identified could be searched for in the overall contingency graph to find other sessions that have similar patterns of activity, to see whether they display similar productivity. Such pattern matching techniques are similar to structural equivalence metrics in social network analysis, which can be employed once contingency graphs are converted into sociograms, as discussed in section 9. Another option is to look outside the session to find influences from or to other sessions. One can trace same-actor and media-dependency contingencies, following the actors and actants respectively. Tracing proceeds forward in time to see whether the new ideas of the session were disseminated elsewhere, or backward in time to identify possible predecessors of the ideational advance. Such an analysis grounds the concept of *brokers* in actual accomplishments, not relying solely on structural relationships that do not guarantee such accomplishments. At this writing we are constructing a contingency graph of several years of data from Tapped-In in preparation for application of methods such as those just described.

7 Uptake Graph: Interaction Model

As discussed above, contingencies are so named because they can include circumstantial relationships between acts with varying degrees of relevance to interaction. Analytic interpretation is required to identify relationships between events that are not merely circumstantial, but reflect intentional acts. An act of *uptake* is one in which an actor takes traces of one or more prior events as having certain significance for an ongoing activity [37]. For example, a speaker takes up some aspect of the prior speaker's utterance, or a message poster in a discussion forum can take up some aspect of the message being replied to. Uptake is a generalization of all interactional relationships used in analysis, such as *comment*, *reply*, *elaboration*. It includes these relationships, but also applies to spatio-temporally distributed associations between actors in which they may not even be aware of each other, let alone be directing their actions towards each other, such as tagging, downloading, etc. Therefore, uptake is more general than *transactivity* [5], which requires other-directedness. Uptake is an appropriate generalized unit of interaction in networked learning environments, where individuals may benefit from each others' presence without conversing directly. The essential idea is that the trace an actor's actions have left in the environment (e.g., chat contribution, discussion posting, uploaded file, profile, recommendation) is taken up by another actor in some manner. Uptake of traces can result in stigmergic effects, i.e., implicit distributed coordination of collective action [30]. Illuminating these stigmergic effects reveals the contingencies

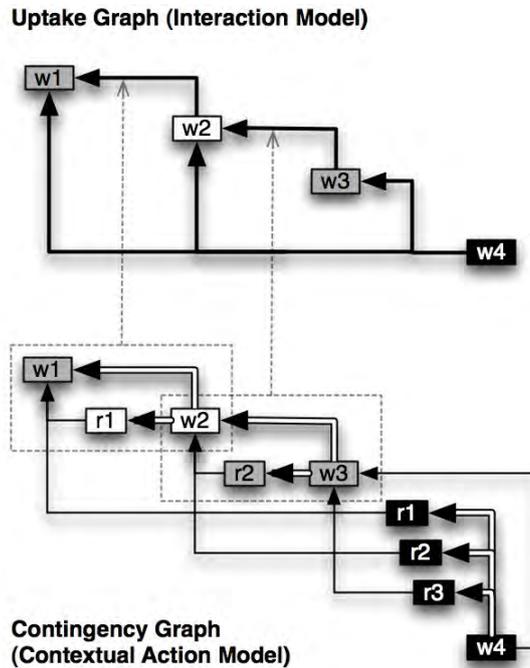


Figure 3. From Contingency to Uptake Graphs

by which an individual's actions are connected to information, actions, and resources from sources that may otherwise not be known to that individual, even if embedded within one's known social network.

An uptake graph is an *interaction model*, as it describes the interaction that the analysis claims is taking place. Although all analytic artifacts from process traces on up involve theoretical decisions, the move from contingency graphs to uptake graphs is a move from primarily empirically accountable representations to those more strongly determined by analytic interpretations. Representationally, an uptake relation is a subgraph of contingencies, as illustrated in Figure 3. An analyst collects contingencies that are considered to be analytically meaningful: a number of contingencies between two or more acts may corroborate the interpretation that the final act is an intentional taking up of traces of the prior ones. For example, *w2*, in which P1 posts a reply to the message posted by P2 in *w1*, is contingent on *w1* in these ways: there is a media dependency (*m2* is linked by threading to *m1*); lexical overlap (*m2* contains phrases also found in *m1*); and a chain of temporal proximity (*w2* took place shortly after read event *r1* by the same actor, and *r1* is media-dependent on *w1* by virtue of reading *m1*). All of these contingencies are taken as evidence for an intentional relationship of *w2* to *w1*, and collapsed into one uptake arc. Because of this relationship between contingencies and uptake, an uptake graph may be seen as an abstraction of a contingency graph, and many of the same analytic moves (such as pattern matching and tracing actions) apply to both.

Contingency and uptake graphs are described more fully in [37]. We have used contingency and uptake graphs to provide interactional accounts of specific accomplishments of participants [24, 25], to trace out information sharing [40], and to detect roles of participants not visible in the final media trace [38]. For example, examining only reply structure (the threading relationship between messages in Figure 1) we might miss the fact that **m4** played an integrative role in this discussion. The uptake graph of Figure 3 makes this integration explicit as a structure of uptake converging on **w4**. Integrative or convergent acts are important to group learning processes such as intersubjective meaning making [35] and community knowledge building [16].

Contingency and uptake graphs represent process models: they focus on how acts relate to each other and constitute a process of interaction. Their basic unit is acts and other events: the actors and entities through which interaction takes place are attributes of these events. Now we turn to an alternative derived representation that makes these actors and entities explicit, rather than the events.

8 Associograms: Mediation Model

In the study of socio-technical networks, we are interested in how the technological infrastructure enables and is utilized by the social actors to interact with each other. The next layer of the analytic hierarchy makes the objects of this technological infrastructure explicit and shows how they mediate interaction between participants. Analysis at this layer provides the *mediation model*, and is represented by multimodal bipartite graphs in which participants are related to each other via the objects through which they interact. We call these graphs *associograms* to distinguish them from sociograms in a manner that honors Latour's [22] concept of mediated associations that assemble a social system. Associograms are multimodal because there may be two or more types of nodes—actors and the various types of media through which they interact—and they are bipartite because they are divided into two partitions: actors in one partition and the various types of media objects in the other. Directed arcs represent state-influence (a weaker form of state-dependency): they extend from an object to an actor if the state of the object is influenced by some action of the actor (e.g., writing a message or editing a wiki), and from the actor to the object if the state of the actor has been influenced by accessing the object (e.g., reading a message or wiki)

One can construct associograms from a set of events, whether taken directly from the event model, or events of interest that were selected from the contextualized action or interaction models (contingency graphs or uptake graphs, respectively). A node in any of these models represents an event, and actors and objects are attributes of the node. This is largely reversed in an associogram: actors and objects are nodes, and events are links between nodes. For example, in Figure 4, **w1**—the event of **P2** writing **m1**—becomes a directed association from **m1** to **P2** (**m1**'s state depends on **P1**), and **r1**—the event of **P1** reading **m1**—becomes a directed association from **P1** to **m1**.

Associogram (Mediation Model)

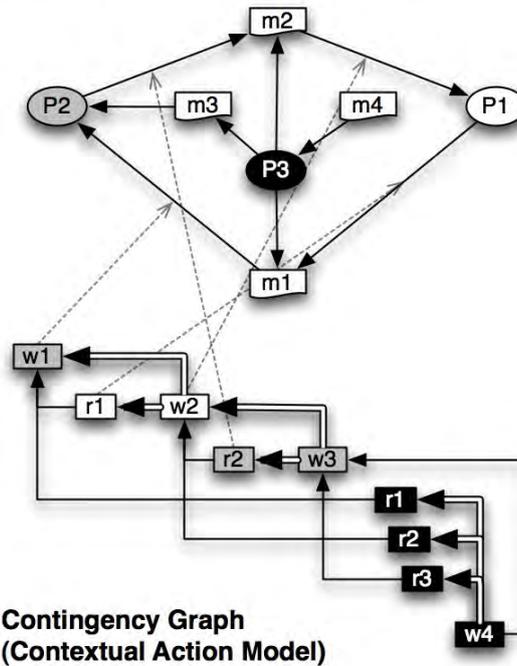


Figure 4. From Events to Associogram

An associogram can be constructed at different granularities. Object nodes could be created for each individual object (e.g., one node for each message, wiki page, chat, etc., as in Figure 4), or they could be aggregated for object types (e.g., all associations via messages aggregated into a single node, those via wikis in another, etc.) in order to characterize how interaction is distributed across types of media. Some information is lost in either case: all the events involving an actor and an object will fall into the same two nodes and links between them. For example, if P1 reads m1 multiple times there is still only one link from P1 to m1, and if P2 edits a wiki multiple times, there is still only one link from the wiki to P2. Some of this information can be preserved by weighting the links with number of occurrences, or by putting backpointers to the originating event nodes. Temporal sequencing is mostly lost, though it can be recovered by following these backpointers to the contingency graph. This information reduction is actually an advantage of associograms: they reduce the clutter of interaction models to expose recurring patterns of mediation. An example is given next.

8.1 Finding Interaction Patterns

Associograms can help expose patterns of interest in contingency or uptake graphs. For example, consider the question of finding which participants are in dialogue with each other. Dialogue is clearly a prerequisite for learning through argumentation, intersubjective meaning-making and group cognition [3, 33, 35]. A key indicator of the presence of dialogue is what we call a *round trip*: one participant makes a contribution that is accessed by another participant who then makes a contingent contribution (evidencing uptake) that the first participant then accesses [40]. In a contingency graph one would need to trace out many paths from each participant to find paths that go to another participant via a read and then a write and then back to the first participant. In an associogram one need only find cycles in the graph. If the links are weighted with frequency counts, the minimum weight of the path is taken as a measure of extent of dialogue. For example, in Figure 4 there is a cycle (following the arrows in reverse to trace chronology rather than dependency) $P2 \leftarrow m1 \leftarrow P1 \leftarrow m2 \leftarrow P2$. This corresponds to the round trip in which P2 posts m1, P1 reads it and posts m2 in reply and P2 reads m2, completing the round trip. Note that P2 need not post a reply to m2 to complete the round trip: an analysis that looks only at the threading structure of posted messages and does not include read events would miss this round trip.

8.2 Characterizing Mediation

Degree and path analysis of an associogram can reveal the roles different media play in a socio-technical network. Media objects or media types (in an aggregate associogram) that have high in-degree are accessed by many actors, and hence may be influential sites where an educational intervention can reach many participants in a socio-technical network. Those with high out-degree are modified by many actors, and hence may be sites where ideas are aggregated or consolidated (potential roles as community memory, or locus of knowledge building). In a weighted associogram, heavily weighted links indicate that actors visit the incident objects repeatedly. These measures may be compared between different media types to assess their relative roles. Additional roles can be identified, such as liaison roles, where the media object or type connects other objects or actors that would not otherwise be reachable. For example, we have used associograms constructed from bridging events to assess the roles of different media (discussions, wikis, resources and profiles) in mediating bridging in a socio-technical network [36].

8.3 Characterizing Mediated Relationships

Associograms summarize how objects directionally mediate the interaction between any given two people. The subgraph of all paths of length two (direct mediation) between two persons can be used in at least two ways to characterize the relationships between those persons as mediated by the socio-technical network. First, we can recognize defined patterns, two of which are shown in Figure 5. Second, profiles of

mediated interaction between any two people can be represented as vectors of the weights on paths of different types and directions (e.g., P1 to P2 via discussions, P2 to P1 via discussions, P1 to P2 via wikis, etc.). Cluster analysis of these vectors can reveal recurring types of relationships. These approaches are currently being investigated in a dissertation by Kar-Hai Chu, under the authors' direction.

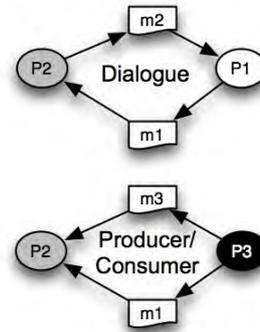


Figure 5. Pairwise Associations (Relationship Model)

9 Sociograms: Tie Model

Finally, we briefly note that associograms can be transformed to conventional sociograms by transitive closure of the paths between actors, or by other computations that interpret patterns of mediated associations as ties. As shown in Figure 6, this results in a directed graph or an asymmetric matrix representing the ties between actors. Well established methods of social network analysis (SNA) can then be applied [41], but with advantages that would not be realized if one had merely constructed sociograms directly from source data (e.g., surveys about ties). A *tie* in a sociogram or sociomatrix is really shorthand for a complex network of multi-mediated interactions that develop over time. If suitable back-pointers to prior representations (the associograms and, via them, the contingency graph) are maintained, then results obtained via network analysis of ties can then be interpreted and understood by expanding back to the mediation and interaction models underlying those ties.

In fact, bidirectional construction and deconstruction of ties was one motivation for the development of this analytic hierarchy. We wanted to leverage the power of SNA for describing patterns of relationships between actors in terms of the structural positions they occupy in relational networks, but wanted to retain the complex and artifact-mediated interactions that ties summarize. In classic SNA research, ties are identified through manual techniques such as interviews or questionnaires, eliciting subjective perceptions of relatedness to others. Some computer-mediated environments allow for a more automated data collection, such as email networks, but those generally represent explicit, intentional communication. Online environments, through activity logging, offer the potential of automatically computing ties from traces of behavior, but they log actions rather than relationships. Therefore, methods to convert *activity* and *interaction* into *ties* are of value. These traces may not reflect subjective perceptions of relationships between persons, but have the advantage that this conversion can be automated. The analytic hierarchy supports such an automated translation from activity logs to mediated interaction and ultimately ties to which SNA methods can be applied. But the value of the analytic hierarchy does not only lie in automating the gathering of tie data for SNA. Analysis and interpretation can go in the other way as well. Once analytic results are reached using summary representations of ties—or using bipartite representations of artifact-mediated

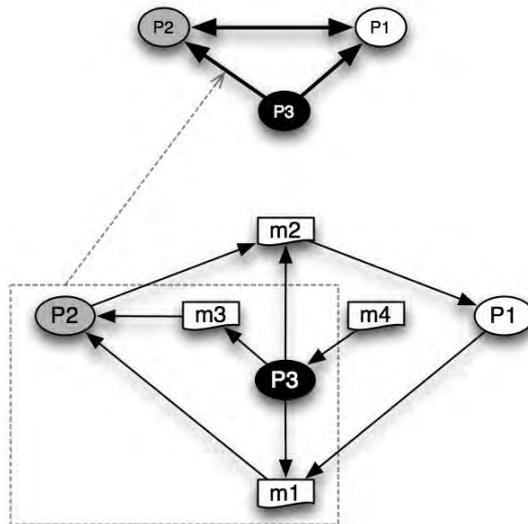


Figure 6. From Associogram to Sociogram

relationships—these results can be interpreted by unpacking the binary ties back into the interactional sequence that they summarize.

Part of the unique power of social network analysis is in the ability to operationalize systemic processes, providing a window into social systems that is otherwise difficult to gain. Taking a systemic approach to learning frames the learning process as a group of individuals that, when interacting with each other, or artifacts of interactive activity, create a whole greater than any isolated individual [9]. If we are to analytically study learning processes it is helpful to have frameworks for analysis that can trace the patterns of interaction generated from learning processes, and harness the information available from the structure revealed by those patterns. The framework presented above provides a missing piece in the analytic ability to extract structural indicators (traces) of activity, both intentionally (e.g. discussion board posts), and unintentionally cast off (e.g. accessing a digital artifact), from individuals interacting and learning in mediated environments.

10 Discussion

10.1 Status

The framework discussed in this paper has been used extensively in our research through mixtures of manual and ad-hoc automated analysis, leading to implementation of a principled analytic software toolkit. Research in our own laboratory is diverse: we study technology-mediated interaction ranging from dyads to online communities, and our methods draw on experimental, conversation-analytic and network analytic traditions. We developed the framework to make distributed

interaction visible, but also as a means of coordinating the various strands of our own work. Initially we developed contingency and uptake graphs through extensive use in manual analyses [24, 38]. Then we implemented software tools for constructing contingency graphs automatically from log files and developed rudimentary analytic tools that leverage these representations [25, 26]; and we have used associograms in ad-hoc computational analyses [36]. More recently we developed a comprehensive “Traces” analytic toolkit that is implemented in Java with Hibernate persistence and is applicable to a variety of socio-technical networks. The Traces design includes: (1) an Entity-Event-Contingency (EEC) core, supporting the fundamental classes of Entity, Event, and Contingency; and classes that are likely to be common to all analyses (e.g., specializations of Entity into ArtifactBase, ActorBase and IdeaBase abstract classes); (2) an Analytic Core layer, including classes for Uptake, various types of Contingencies, and Composites (CompositeEntity, e.g., discussion forums; CompositeEvent, e.g., chat sessions; and CompositeContingency, used to represent uptake); (3) a Domain Core, with common domain objects such as Chat and Contribution, Discussion and Message, etc. These cores are extended for application to specific data sources such as Tapped-In, using (4) data source specific extensions to the Domain Core such as Calendar Events (these may migrate to the core if they are found to be useful across systems, and (5) classes that map the data source databases and logs to the above. Presently the Traces toolkit does not include associograms: our analyses at the levels of mediation and tie models are handled by exporting to other tools available for the task at hand (e.g., Jung and UCINet). We have imported two years of data from Tapped-In, and are conducting analyses to be reported in future publications.

10.2 Multi-Level Analysis and Theoretical Multivocality

Although this paper outlines how the analytic hierarchy is constructed as one goes from log data to more abstract representations of action, interaction, mediation and tie, it should be emphasized that the analytic hierarchy is not just a data interpretation framework. It is also intended to be a structural framework for connecting theorizing at different levels. Developmentally, the framework arose out of our own need to reconcile our research on small group interaction in computer supported collaborative learning and online learning contexts with our emerging research on online communities. It was clear that studies of communication networks, and social network analysis in particular, had something to offer, but the “ties” of such analyses seemed to hide away the very processes we were interested in: the interaction and how it was influenced by and appropriated the media we were designing. Therefore we constructed the framework as an explicit bridge between analyses of local interaction and of larger social phenomena, with the expectation that it would also guide our bridging between theoretical explanations at these different levels.

This work is sympathetic to calls by Contractor and colleagues [7, 27] for multi-theoretical and multi-level (MTML) analyses and models of communication networks, but is based on a different conception of layering than MTML. The MTML approach calls for examining (1) the properties of individual nodes (incorporating attributed-based data); (2) properties of the network under consideration (including

dyadic, triadic and global properties); and (3) relationships of this network to other relations over the same network constituents or the same relations as they change over time. These levels change the granularity or scope of analysis, but stay within an ontology of structural relations between a set of network constituents. Our hierarchical approach adds a more “vertical” dimension, changing the ontology between layers, from relationships between observed events, to mediated associations, to direct ties between actors.

Marin & Wellman [23] contrast attribute based explanations, which explain behavior in terms of attributes of individuals, with the network analysis' position “that causation is not located in the individual, but in the social structure.” Similarity of attributes is explained by similarity of network positions, due to the similar “constraints, opportunities and perceptions created by these similar network positions.” We agree with their critique of attribute-based explanations but wish to avoid the opposite oversimplification: individuals' similarities do not arise merely out of static structures piping influences into the individuals from without. A range of thinkers, including Garfinkel [13], Blumer [6] and Latour [22] have argued (each in their own way) that social regularities are constructed and sustained through interaction between actors (whether strictly local interaction, as for Garfinkel, or potentially mediated across time and space, as for Latour). To fully understand social systems we must examine interaction. Colleagues² have offered the analogy that ethnomethodological interaction is the quantum mechanics of social science. We can ignore it when explaining social life at a Newtonian level, but to really understand the origins of the social world we must dive in and find how fluctuations in micro-phenomena can have an influence on larger scale change. Latour [22] has made a similar observation in claiming that the “sociology of the social” may seem adequate for explanation of stable states of affairs, but Actor-Network Theory's “sociology of associations” is needed to understand rapidly changing networks.

Our position is that network structures are relevant because of how they support interaction. The network structure is not enough: to explain the origins of social life we must understand the nature of the communication or interaction that takes place. In socio-technical networks, this includes understanding how that interaction is embedded in and exploits the resources of technological infrastructures; i.e., how it is *mediated*. The present work offers a conceptualization of how to map between these different levels of theory and analysis, viz., structure, mediation, and interaction; and also provides specific representations for supporting analytic work with computational tools.

In addition to providing a unified analytic artifact and supporting multiple levels of analysis, a third concern has motivated the work reported here. Researchers from multiple theoretical and analytic traditions are studying distributed and networked learning, virtual organizations, and similar socio-technical networks. This diversity can mean balkanization, or it can be a strength. A single integrated discipline of Learning Analytics may not be possible or even desirable, but there must be some basis for dialogue and coordination between the traditions. Shared instruments and representations mediate the daily work of scientific discourse [21], and advances in

² Ravi Vatrapu, personal communication, July 28, 2007; David Sallach, personal communication, May 23, 2010

scientific disciplines are sometimes accompanied with representational advances. Similarly, researchers studying socio-technical networks could benefit from shared ways of conceptualizing and representing distributed interaction, or at least from boundary objects [34] that make discourse between multiple analytic traditions possible. We offer this framework as a potential basis productive discourse among multiple analytic voices [39] in the study of socio-technical networks such as networked learning by enabling the development of shared conceptualizations, representations, and tools at a given level of analysis and supporting bridging between different levels of analysis.

Acknowledgements

This work was supported by NSF Award #0943147. The view expressed herein do not necessarily represent the views of NSF. The authors thank Nathan Dwyer, Richard Medina, Ravi Vatrapsu, and Kar-Hai Chu for discussions and collaborations through which the present framework was refined.

References

- [1] I. E. Allen and J. Seaman, *Growing by Degrees: Online Education in the United States, 2005*, Alfred P. Sloan Foundation, Needham, MA, 2005.
- [2] T. Anderson, ed., *The Theory and Practice of Online Learning, Second Edition*, Athabasca University Press, Edmonton, Canada, 2008.
- [3] J. Andriessen, M. Baker and D. Suthers, *Argumentation, Computer Support, and the Educational Context of Confronting Cognitions*, in J. Andriessen, M. Baker and D. Suthers, eds., *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003, pp. 1-25.
- [4] S. A. Barab, R. Kling and J. H. Gray, *Designing for Virtual Communities in the Service of Learning*, Cambridge University Press, New York, 2004.
- [5] M. W. Berkowitz and J. C. Gibbs, *A Preliminary Manual for Coding Transactive Features of Dyadic Discussion (unpublished manual)*, Ohio State University, 1979.
- [6] H. Blumer, *Symbolic Interactionism: Perspective and Method*, University of California Press, Los Angeles, 1986.
- [7] N. S. Contractor, S. Wasserman and K. Faust, *Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example*, *Academy of Management Review*, 31 (2006), pp. 681-703.
- [8] J. N. Cummings, T. Finholt, I. Foster, C. Kesselman and K. A. Lawrence, *Beyond Being There: A Blueprint for Advancing the Design, Development and Evaluation of Virtual Organizations (Final report from workshops on building effective virtual organizations)* 2008.
- [9] M. de Laat, V. Lally, L. Lipponen and R.-J. Simons, *Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis*, *International Journal of Computer Supported Collaborative Learning*, 2 (2007), pp. 87-103.
- [10] S. J. Derry and G. Fischer, *Transdisciplinary Graduate Education*, *American Educational Research Association*, Montreal, Canada, 2005.

- [11] A. Duranti, *Transcripts, Like Shadows on a Wall*, Mind, Culture & Activity, 13 (2006), pp. 301-310.
- [12] U. Farooq, P. Schank, A. Harris, J. Fusco and M. Schlager, *Sustaining a community computing infrastructure for online teacher professional development: A Case Study of Designing Tapped In*, Computer Supported Cooperative Work, 16 (2007), pp. 397-429.
- [13] H. Garfinkel, *Studies in Ethnomethodology*, Prentice-Hall, Englewood Cliffs, New Jersey, 1967.
- [14] D. R. Garrison and H. Kanuka, *Blended learning: Uncovering its transformative potential in higher education*, The Internet and Higher Education, 7 (2004), pp. 95-105.
- [15] R. Goldman, R. Pea, B. Barron and S. J. Derry, *Video Research in the Learning Sciences*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2007.
- [16] J. Hewitt, *Beyond threaded discourse*, International Journal of Educational Telecommunications, 7 (2001), pp. 207-221.
- [17] B. Jordan and A. Henderson, *Interaction Analysis: Foundations and practice*, The Journal of the Learning Sciences, 4 (1995), pp. 39-103.
- [18] S. Joseph, V. Lid and D. D. Suthers, *Transcendent Communities*, in C. Chinn, G. Erkens and S. Puntambekar, eds., *The Computer Supported Collaborative Learning (CSCL) Conference 2007*, International Society of the Learning Sciences, New Brunswick, 2007, pp. 317-319.
- [19] R. Kling, *Learning about information technologies and social change: The contribution of social informatics*, The Information Society, 16 (2000), pp. 217-232.
- [20] K. Koedinger, D. D. Suthers and K. Forbus, *Component-based construction of a science learning space*, International Journal of Artificial Intelligence in Education, 10 (1999), pp. 292-313.
- [21] B. Latour, *Drawing things together*, in M. Lynch and S. Woolgar, eds., *Representation in Scientific Practice*, MIT Press, Cambridge, MA, 1990, pp. 19-68.
- [22] B. Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford University Press, New York, 2005.
- [23] A. Marin and B. Wellman, *Social Network Analysis: An Introduction*, in P. Carrington and J. Scott, eds., *Handbook of Social Network Analysis*, Sage, London, 2010.
- [24] R. Medina, D. Suthers and R. Vatrappu, *Inscriptions becoming representations*, in C. O'Malley, P. Reimann, D. Suthers and A. Dimitracopoulou, eds., *Computer Supported Collaborative Learning Practices: CSCL 2009 Conference Proceedings*, International Society of the Learning Sciences, Rhodes, Greece, 2009, pp. 18-27.
- [25] R. Medina and D. D. Suthers, *Bringing Representational Practice From Log to Light*, in P. A. Kirschner, F. Prins, V. Jonker and G. Kanselaar, eds., *International Perspectives in the Learning Sciences: Creating a Learning World: Proceedings of the Eighth International Conference for the Learning Sciences (ICLS 2008)*, International Society of the Learning Sciences Utrecht, 2008, pp. 59-66.
- [26] R. Medina and D. D. Suthers, *Using a contingency graph to discover representational practices in an online collaborative environment*, Research and Practice in Technology Enhanced Learning, 4 (2009), pp. 281-305.
- [27] P. R. Monge and N. S. Contractor, *Theories of Communication Networks*, Oxford University Press, Oxford, 2003.
- [28] E. Ochs, *Transcription as theory*, in E. Ochs and B. B. Schieffelin, eds., *Developmental Pragmatics*, Academic Press, New York, 1979, pp. 43-72.
- [29] K. A. Renninger and W. Shumar, *Building Virtual Communities: Learning and Change in Cyberspace*, Cambridge University Press, Cambridge, 2002.

- [30] D. Rosen and D. D. Suthers, *Stigmergy and collaboration: Tracing the contingencies of mediated interaction*, *Proceedings of the Hawaii International Conference on the System Sciences (HICSS-44), January 4-7, 2011, Kauai, Hawai'i (CD-ROM)*, Institute of Electrical and Electronics Engineers, Inc. (IEEE), New Brunswick, 2011.
- [31] P. Sanderson and C. Fisher, *Exploratory sequential data analysis: Foundations*, *Human-Computer Interaction*, 9 (1994), pp. 251-318.
- [32] M. Schlager, J. Fusco and P. Schank, *Evolution of an Online Education Community of Practice*, in K. Renninger and W. Shumar, eds., Cambridge University Press, *Building Virtual Communities*, 2002, pp. 129-158.
- [33] G. Stahl, *Group Cognition: Computer Support for Collaborative Knowledge Building*, MIT Press, Cambridge, MA, 2006.
- [34] S. L. Star and J. R. Griesemer, *Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology*, *Social Studies of Science*, 19 (1989), pp. 387-420.
- [35] D. D. Suthers, *Technology affordances for intersubjective meaning-making: A research agenda for CSCL*, *International Journal of Computer Supported Collaborative Learning*, 1 (2006), pp. 315-337.
- [36] D. D. Suthers and K.-H. Chu, *Identifying mediators of socio-technical capital in a networked learning environment*, in L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell and T. Ryberg, eds., *Proceedings of the 7th International Conference on Networked Learning*, Aalborg, Denmark, 2010, pp. 387-395.
- [37] D. D. Suthers, N. Dwyer, R. Medina and R. Vatrappu, *A framework for conceptualizing, representing, and analyzing distributed interaction*, *International Journal of Computer Supported Collaborative Learning*, 5 (2010), pp. 5-42.
- [38] D. D. Suthers, N. Dwyer, R. Vatrappu and R. Medina, *Analyzing Interactional Construction of Meaning in Online Learning*, *Proceedings of the 39th Hawai'i International Conference on the System Sciences (HICSS-39), January 4-7, 2006, Poipu, Kauai (CD-ROM)*, IEEE Computer Society Press, Poipu, Hawai'i, 2006.
- [39] D. D. Suthers, K. Lund, C. Rosé, G. Dyke, N. Law, C. Teplovs, W. Chen, M. M. Chiu, H. Jeong, C.-K. Looi, R. Medina, J. Oshima, K. Sawyer, H. Shirouzu, J.-W. Strijbos, S. Trausan-Matu and J. van Aalst, *Towards productive multivocality in the analysis of collaborative learning*, in H. Spada, G. Stahl, N. Miyake, N. Law and K. M. Cheng, eds., *Connecting Computer-Supported Collaborative Learning to Policy and Practice: Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning (CSCL 2011)* University of Hong Kong, Hong Kong, in press.
- [40] D. D. Suthers, R. Medina, R. Vatrappu and N. Dwyer, *Information sharing is incongruous with collaborative convergence: The case for interaction*, in C. Chinn, G. Erkens and S. Puntambekar, eds., *The Computer Supported Collaborative Learning (CSCL) Conference 2007*, International Society of the Learning Sciences, New Brunswick, 2007, pp. 714-716.
- [41] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, 1994.
- [42] E. Wenger, R. A. McDermott and W. Snyder, *Cultivating Communities of Practice: A Guide to Managing Knowledge*, Harvard Business School Press, Boston, Mass., 2002.