

Cohesion network analysis of CSCL participation

Mihai Dascalu^{1,2}  · Danielle S. McNamara² · Stefan Trausan-Matu¹ · Laura K. Allen²

Published online: 13 April 2017
© Psychonomic Society, Inc. 2017

Abstract The broad use of computer-supported collaborative-learning (CSCL) environments (e.g., instant messenger–chats, forums, blogs in online communities, and massive open online courses) calls for automated tools to support tutors in the time-consuming process of analyzing collaborative conversations. In this article, the authors propose and validate the cohesion network analysis (CNA) model, housed within the ReaderBench platform. CNA, grounded in theories of cohesion, dialogism, and polyphony, is similar to social network analysis (SNA), but it also considers text content and discourse structure and, uniquely, uses automated cohesion indices to generate the underlying discourse representation. Thus, CNA enhances the power of SNA by explicitly considering semantic cohesion while modeling interactions between participants. The primary purpose of this article is to describe CNA analysis and to provide a proof of concept, by using ten chat conversations in which multiple participants debated the advantages of CSCL technologies. Each participant’s contributions were human-scored on the basis of their relevance in terms of covering the central concepts of the conversation. SNA metrics, applied to the CNA sociogram, were then used to assess the quality of each member’s degree of participation. The results revealed that the CNA indices were strongly correlated to the human evaluations of the conversations. Furthermore, a stepwise regression analysis indicated that the CNA indices collectively predicted 54% of the variance in the human ratings of participation. The results provide promising support for the use of automated computational

assessments of collaborative participation and of individuals’ degrees of active involvement in CSCL environments.

Keywords Cohesion network analysis · Computer-supported collaborative learning · Cohesion-based discourse analysis · Participation evaluation · Dialogism · Polyphonic model

The rise in modern technological advances has resulted in dramatic shifts in education: how education is delivered, who participates in educational activities, and how students interact with their instructors and each other. Increasing numbers of individuals actively seek educational opportunities, as well as opportunities to interact with other, to share their opinions, and to collaborate online. Technology facilitates access to knowledge and learning, which can now be achieved without locational boundaries. Following the increasing popularity of e-learning and computer-aided remote education, a recent educational framework has gained momentum—computer-supported collaborative learning (CSCL; Stahl, Koschmann, & Suthers, 2006). CSCL promotes collaboration in digital learning contexts through dedicated learning platforms on which groups of students can share, discuss, and exchange ideas (Stahl, 2006), thus empowering traditional learning methods with shared expertise (Cress, 2013). The switch from traditional educational systems to collaborative environments facilitated by CSCL technologies strengthens the bonds between learners through information sharing and open discussions. In addition, the enhancement of individual learning with collaboration from CSCL sessions, which engage learners in open discussions, is particularly beneficial for problem-solving tasks (Stahl, 2006). For these reasons, CSCL has emerged as a viable educational option that enables learners worldwide to gain access to information and exchange expertise, potentially reducing educational gaps

✉ Mihai Dascalu
mihai.dascalu@cs.pub.ro

¹ University Politehnica of Bucharest, Bucharest, Romania

² Arizona State University, Tempe, AZ, USA

between learners that can emerge as a function of physical distance or culture.

As an educational framework, CSCL currently relies on the availability of controlled learning environments that are monitored by human tutors. Indicators such as performance, participation, or collaboration with other members are useful when measured at the level of the individual in order to ensure an effective learning process for each participant. However, such evaluation is a time-consuming process and instructors are faced with the problem of not having sufficient time to administer and score deep and meaningful measures of student participation and collaboration. Indeed, this is one important drawback of CSCL (Trausan-Matu, 2010a). Therefore, automated tools that enable monitoring and scoring are necessary for CSCL to be successfully applied at larger scales.

The aim of this article is to propose and validate an automated *cohesion network analysis* (CNA) model to describe and evaluate participation in CSCL conversations. This model relies on *dialogism* as a theoretical background and on *cohesion* as the underlying discourse structure (both described in detail in the following subsections). This model goes far beyond previously proposed models for automated assessment, which rely solely on counting the number of utterances exchanged between different speakers. Our automated participation model is housed within ReaderBench (Dascalu, Stavarache, Dessus, et al. 2015; Dascalu, Stavarache, Trausan-Matu, et al. 2015), a fully functional automated software framework, designed to provide support for students and tutors through assessments and predictions of comprehension in various educational contexts. The system makes use of text mining techniques based on advanced natural language processing (NLP) and machine learning algorithms to design and deliver summative and formative assessments using multiple data sets (e.g., CSCL conversations, online community discussions, assigned textual materials, students' self-explanations). The quantitative indices introduced by our CNA participation model are extensible and can be used to assess participation in different collaborative groups, such as academic chats, course forums, and online knowledge-building communities of practice.

This study builds upon the authors' previous work on discourse cohesion (Dascalu, Trausan-Matu, & Dessus, 2013; McNamara, Graesser, McCarthy, & Cai, 2014; Trausan-Matu, Dascalu, & Dessus, 2012) by performing an in-depth analysis of participation in CSCL contexts. In contrast to previous studies (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015a, b; Dascalu, Trausan-Matu, McNamara, & Dessus, 2015), which have focused on the assessment of collaboration, the present study shifts the perspective toward a complementary dimension of CSCL—participation. Importantly, the CNA model introduced in this article is flexible: it can be used to provide in-depth, discourse-centered assessments of participation, active involvement, and engagement in any CSCL environment that makes use of learners' text productions.

Dialogism and the polyphonic model

One of the most important components of CSCL is that learning can be seen as a collaborative knowledge-building process (Bereiter, 2002; Scardamalia & Bereiter, 2006). Small groups of students interact (Stahl, 2006) and inter-animate in a polyphonic way (Trausan-Matu, Stahl, & Sarmiento, 2007), rather than participate in knowledge transfer from the teacher to the learner. Moreover, if students receive tasks in their zone of proximal development (ZPD; Vygotsky, 1978), the learning process may be seen as having two intertwining cycles: a personal one and a social knowledge building one (Stahl, 2006).

Dialogism is considered by many as a viable theoretical framework for CSCL in which discourse is modeled as the interaction with others, oriented toward building meaning and understanding (Arnseth & Ludvigsen, 2006; Koschmann, 1999; Stahl, Cress, Ludvigsen, & Law, 2014; Stahl et al., 2006; Trausan-Matu, Stahl, & Sarmiento, 2007; Wegerif, 2005). The idea of *dialogism* was introduced by Mikhail Bakhtin (Bakhtin, 1981, 1984) and covers a broader, more abstract, and more comprehensive sense of dialogue that is reflected in any of the following perspectives: communication, interaction, action, or cognitive process (Linell, 2009, pp. 5–6). This definition of dialogism, besides the intrinsic dialogue between different individuals, may be present in any kind of text—since life is dialogic by its very nature (Bakhtin, 1984, p. 294). In addition, dialogue can be also perceived as an “internal dialogue within the self” or an “internal dialogue” (Linell, 2009, ch. 6), a “dialogical exploration of the environment” (Linell, 2009, ch. 7), a “dialogue with artifacts” (Linell, 2009, ch. 16) or a “dialogue between ideas” (Marková, Linell, Grossen, & Salazar Orvig, 2007, ch. 6).

In each context, discourse is modeled from a dialogical perspective as interaction with others, essentially toward building meaning and understanding. In other words, the dialogical framework is centered on sense-making (Linell, 2009), with emphases on:

- a) *action*: Wertsch (1998) suggests that actions are the building blocks of the mind, and meaning is constructed through interactions with others and the world in a given context;
- b) *cognition*: we acquire knowledge about the world and assign meaning to it through language and interaction within a specific context; and
- c) *communication*: the interaction with others generates the meaning of discourse and also incorporates a strong cognitive component as “every authentic function of the human spirit [...] embodies an original, formative power” (Cassirer, 1953, p. 78).

From a broader point of view, *discourse* is defined in NLP as “a coherent structured group of sentences” (Jurafsky &

Martin, 2009, ch. 21) and has different connotations for monologues and dialogues, which rely on either uni- or bidirectional communications (Trausan-Matu & Rebedea, 2010). Monologues are characterized as one-way, speaker–listener-directed communication models (Jurafsky & Martin, 2009). For these models, the usual manner of analyzing discourse consists of segmenting texts, identifying different relationships among text segments, and analyzing the cohesion or coherence between their ideas (McNamara et al., 2014).

In terms of discourse analysis of coherence relations, probably the most well-known theories were proposed by Hobbs (1985), Grosz, Weinstein, and Joshi (1995), and Mann and Thompson (1987). Hobbs' theory is built on semantic coherence relations between the current utterance and the preceding discourse (Hobbs, 1978, p. 2) and on abduction inferences in formal logic (Hobbs, 1979, 1985). Rhetorical structure theory (RST; Mann & Thompson, 1987) uses hierarchical rhetorical structures between text spans (i.e., contiguous intervals of text) that are classified as nuclei or satellites in accordance to their importance. These links between text spans are built using a set of rhetorical schemas (patterns) out of which the most frequently used are: antithesis and concession, enablement and motivation, interpretation and evaluation, restatement and summary, and elaboration. In contrast, centering theory (Grosz, Weinstein, & Joshi, 1995) reflects coherence at both local (coherence among the utterances in a given segment) and global levels (coherence with other segments of the discourse) by considering two types of centers (backward-looking and forward-looking) encountered in the intentional and attentional states.

These discourse theories, although useful when applied to texts or monologues, are not directly applicable to dialogue analysis. Their adequacy is primarily limited by the mixture of utterances of more than two speakers and the inter-twining of different conversation threads, which is frequent in CSCL chat conversations. Moreover, as Hobbs (1990) observed, the phenomenon of topic drifting is frequently encountered in spoken conversations due to three occurring mechanisms: semantic parallelism, chained explanations, and metatalk. Although adjacent segments are coherent, the end of the conversation can be significantly different from its starting point.

Our aim is to use a more generalizable model that can be more easily applied to multi-participant conversations. The *polyphonic theory* of CSCL (Trausan-Matu, 2010b; Trausan-Matu & Rebedea, 2009; Trausan-Matu, Rebedea, & Dascalu, 2010; Trausan-Matu, Stahl, & Zemel, 2005) follows the ideas of Koschmann (1999) and Wegerif (2005), and investigates how Bakhtin's dialogism theory, centered on polyphony and inter-animation (Bakhtin, 1981, 1984), can be used to analyze such conversations. Other attempts to analyze conversations with multiple participants have considered other global perspectives, such as *transactivity*, which focuses on argument sequences and how learners build upon their learning partners'

contributions (Joshi & Rosé, 2007). However, most of these perspectives are also based on the two interlocutors model (Trausan-Matu & Rebedea, 2010) mentioned above, and are not easily applicable to all kinds of CSCL conversations, ranging from chats to forums and blogs. Thus, a differentiator of the polyphonic model is its capability to consider and capture the intertwining of different conversation threads, a dimension that is not modeled well by the previously presented models.

The polyphonic model of discourse (Trausan-Matu, 2010b) may be applied to more than two interlocutors and it is a natural way of assessing participation considering both instant, transversal interacting actions, and threads of actions. A participant in a polyphonic framework is participating with individual contributions oriented, however, toward achieving a coherent whole through collaboration, in a similar way to polyphonic music. In conversations, this means that the individual should emit semantically significant utterances to the joint discourse and, meanwhile, consistently interact with others in the conversation. Thus, we can capture multiple simultaneous discourse structures that inter-animate and create a polyphonic weaving. These notions provide the foundation of our method, which is presented in the following section on cohesion network analysis.

Discourse cohesion

The term *cohesion* refers to the incidence of explicit lexical, grammatical, or semantic text cues that help readers make connections among the presented ideas. Halliday and Hasan (1976) provided a detailed analysis of cohesion and suggested that it can be represented as the “relations of meaning that exist within the text, and that define it as a text” (Halliday & Hasan, 1976, p. 4). In analyses of discourse, cohesion plays an important role in identifying the structural relations between the main components of discourse (McNamara, Louwerse, McCarthy, & Graesser, 2010). Multiple approaches can be used to assess textual cohesion, including the frequency of discourse connectors such as cue words (e.g., “but,” “because”) or phrases (e.g., “in order to,” “on the other hand”; McNamara et al., 2014), referring expressions (e.g., nouns that function to identify some object or event; Jurafsky & Martin, 2009), and the semantic similarity between concepts in the text. Semantic similarity can be represented in a number of ways, such as through the semantic “distance” calculated between words in lexical networks (Budanitsky & Hirst, 2006) or through the use of semantic models such as latent semantic analysis (LSA; Landauer & Dumais, 1997), and latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003), which are described in detail in the following section. These semantic distance models have been used to assess text cohesion in isolation, as well as in combination with other textual metrics, such as word repetition (Dascalu, 2014).

Within the context of chat conversations, *high cohesion* denotes a consistent discourse among participants in terms of the topics approached (see Table 1), whereas *low cohesion* is typically indicative of topic changes, multiple concurrent discussion threads, or off-topic contributions (see Table 2; Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). Both Tables 1 and 2 present excerpts from the chat conversations used to validate the CNA model introduced later in the article. In addition, the excerpts introduce explicit reference identifiers to previous utterances from the conversation that were introduced by the users while discussing via the ConcerChat (Holmer, Kienle, & Wessner, 2006) graphical user interface. Table 1 has longer contributions with more elaborated ideas centered on CSCL technologies, which can be used together to define a new collaboration framework. Semantically related concepts are frequently used together (e.g., “project”–“company”–“customer”–“employee”–“staff”–“productivity”–“technology”), and the contributions are highly cohesive within the presented context. The excerpt from Table 2 is characterized by frequent shifts between technologies and, although the framing is the same (i.e., the benefits and disadvantages of CSCL technologies), the points of view vary greatly and make references to completely different external concepts, thus decreasing the overall cohesion.

Because of its focus on the connections among text ideas, cohesion is expected to be strongly correlated with the concept of interanimation in dialogism. Dialogism is based on the interanimation of voices viewed in an extended sense (Trausan-Matu, Dascalu, & Rebedea, 2014), or participants’ points of view, which by their very nature and definition are cohesive, including both convergences and divergences (Trausan-Matu, Stahl, & Sarmiento, 2007). Although the act of reaching convergence and consensus is cohesive by its nature, divergence also might create the premises of a cohesive dialogue in which potentially opposite points of view relating to a single given topic are exposed and eventually

may converge to a consensus. In any case, cohesion is the foundation of a dialogical discourse in which different points of view are linked together in a cohesive manner.

Cohesion network analysis

CNA is theoretically grounded in dialogism and relies on cohesion indices to analyze the structure of a particular discourse. As is shown in Fig. 1, the CNA participation model (a) starts from the cohesion graph as an underlying discourse representation, (b) applies the cohesion scoring mechanism, and (c) uses the sociogram to model the quality of the dialogue presented as a multithreaded polyphonic structure, thus generating four quantitative indices to estimate the participation of each CSCL member. These stages are described in detail in the following sections.

Overall, CNA builds a cohesion-centered, macro level representation of discourse by relying on micro-level content or, more specifically, discourse constituents present in the participants’ contributions that comprise the conversation’s discussion threads. Once this representation has been generated, multiple quantitative indices can be extracted and used to predict the degree of participation or overall engagement of learners in CSCL conversations.

Cohesion graph

Our overall aim is to computationally assess learners’ participation in CSCL conversations through the development and application of a computational model of the cohesion of a particular discourse. CNA provides a means to score utterances and analyze discourse structure within collaborative conversations by combining NLP techniques with social network analysis (Newman, 2010; Wasserman & Faust, 1994). To this end, the CNA model first estimates the cohesion of a CSCL conversation through the use of multiple semantic similarity metrics (Dascalu, Trausan-Matu, McNamara, &

Table 1 Conversation excerpt denoting high cohesion between the contributions from multiple participants

Participant ID	Utterance ID	Referenced Utterance ID	Text
1	18		there are many things we can consider: wikis, google wave, forums, blogs, chat and many more
2	19		let’s begin with the description of our project
1	20	19	well, our software company has produced many applications for mobile phones and other mobile devices. so far we have many satisfied customers and employees, but we need more
3	21		basically we need to add some ways for our employees to communicate better, in order to increase our productivity
1	22		one of the essential things we need is good collaboration between our staff members, and that includes everything from chit-chat to technical details
2	23	20	Ok and in order to do this we need to use the best technologies.

Table 2 Conversation sample denoting dialogism, including divergences, but a lower cohesion between adjacent contributions specific to brainstorming sessions

Participant ID	Utterance ID	Referenced Utterance ID	Text
1	39	36	in chats everything can be messy, but the only thing really important is that you get your answer very fast
2	40	29	with forums on your website you can earn a lot of money
3	41	37	the same thing with blogs you can build a community and be in contact one with each other all the time
2	42	40	if the website is well advertised
1	43	38	you are missing something...the success of the Wikipedia may not necessarily be replicated elsewhere.
1	44	38	and, most important, a collaborative Wiki may suffer from a lack of a strong vision or leadership

Dessus, 2015). Figure 2 introduces the overall automated evaluation process whose stages are presented in detail in this section. Our method can be employed on any conversation transcripts or discussion threads generated separately from CSCL environments that afterward become input files in our processing pipeline.

The chat conversations are first preprocessed using specific NLP techniques (Manning & Schütze, 1999), such as tokenization, splitting, part-of-speech tagging, parsing, stop words elimination, stemming, and lemmatization. The cohesion score is then calculated as an aggregate of semantic distances (Budanitsky & Hirst, 2006). In alignment with our previous studies (Dascalu, 2014), we assess cohesion using a combination of techniques, specifically a (nonlatent) word-based index (i.e., the Wu–Palmer ontology-based semantic similarity; Wu & Palmer, 1994), combined with latent semantic analysis (LSA) and latent Dirichlet allocation (LDA), both described in detail within this section. These indices act as complementary components to better reflect semantic relationships in comparison to a single semantic model.

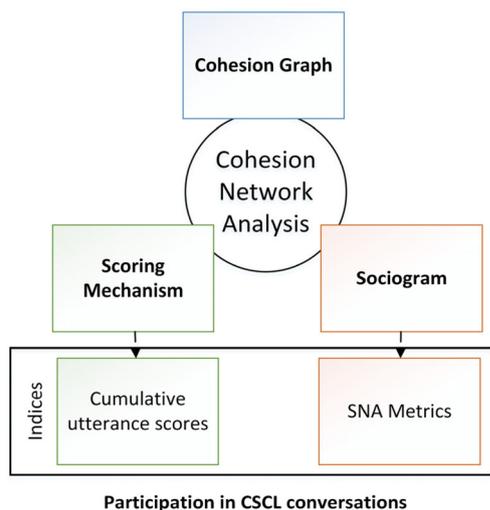


Fig. 1 Visual representation of the CNA model used to assess participation

The cohesion indices introduced in Fig. 1 and described later on in detail are then applied to a social network comprising dialogue in order to estimate the connections between discourse elements. The resulting *cohesion graph* (Dascalu et al., 2013; Trausan-Matu et al., 2012), a generalization of

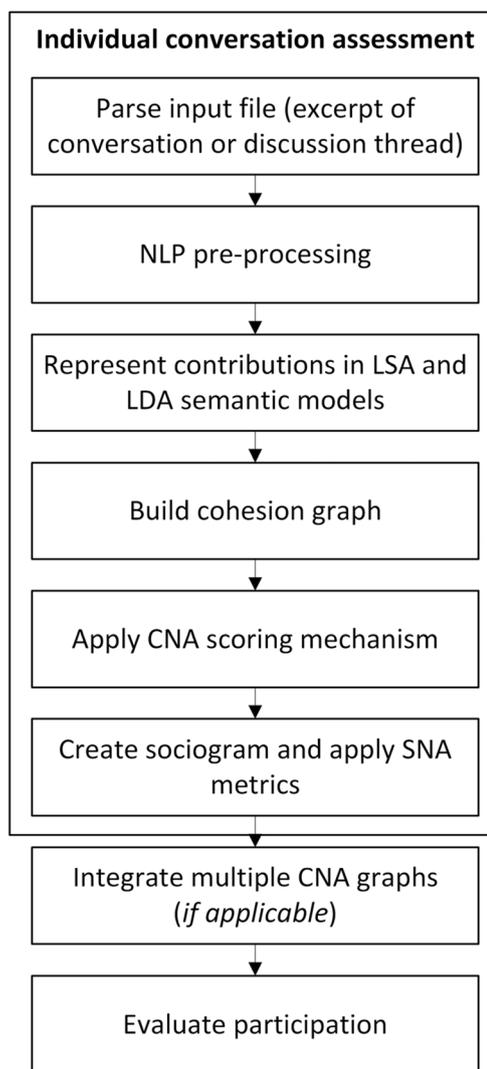


Fig. 2 CNA automated processing workflow

the utterance graph (Trausan-Matu, Stahl, & Sarmiento, 2007), serves as a proxy for the underlying semantic content of the discourse (McNamara et al., 2014).

This cohesion graph is multilayered and contains different types of nodes (Dascalu, 2014). The entire conversation is represented as the central node that is decomposed into participants' contributions and, subsequently, into the underlying sentences and words. Between different layers of the hierarchy, cohesive links are introduced in order to measure the strength of the inclusion, which is represented in terms of the relevance of an utterance with respect to the entire conversation or the impact of a word for each contribution. In addition, mandatory links are established between adjacent utterances in order to model the information flow throughout the discourse. These adjacency links are useful for identifying cohesion gaps that are most likely caused by a change in the discussed topics or a shift towards a different discussion thread. The explicit links, added by the participants within the graphical interface to denote discourse relatedness (see e.g., ConcertChat; Holmer et al., 2006), are also integrated within the cohesion graph as mandatory links. In terms of explicit links, collaborative environments (in our specific case, ConcerChat) allow users to make explicit graphical links to prior contributions, including “reply-to” functionality. Moreover, some CSCL environments include the possibility to share objects in a white board, thus creating a different kind of coherence links. The latter type of links are not considered in our model, which relies solely on text inputs.

In addition, cohesive links are introduced as connectors between possible highly related contributions using an imposed window of 20 utterances. The latter value was experimentally determined as an analogy to the observation that users feel the need to add explicit links to previous utterances within a maximum window of 20 utterances. Specifically, Rebedea (2012) reported that more than 99% of the explicit links created by users in the CSCL chat environment are within a span of 20 utterances, thus arguing for the dimension of our analysis window. Pairwise comparisons are performed for all contribution pairs within the previously defined sliding windows. Links between two selected contributions that have a relatedness value higher than the average plus standard deviation of all pairwise LSA and LDA semantic similarity scores (described later on) are added to our CNA graph as cohesive links.

Semantic similarity with latent semantic analysis LSA is an NLP technique that highlights co-occurrence relations between words and text documents through the development of a vector-space representation of semantic information (Deerwester et al., 1989; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumais, 2004; Landauer & Dumais, 1997). The resulting vector-space model is used to evaluate the semantic similarity between words and text

documents (Landauer, Foltz, & Laham, 1998; Manning & Schütze, 1999). To develop these semantic models, LSA applies an unsupervised learning process to a large corpus of natural language texts, which are relevant for a particular domain. This process first involves the calculation of a sparse term-document matrix that designates the occurrence of individual words in corresponding documents. LSA relies on a “bag-of-words” approach as it disregards word order and uses only normalized term occurrences. The indirect link induced between groups of terms and documents is obtained through a singular-value decomposition (SVD; Golub & Reinsch, 1970; Landauer, Laham, & Foltz, 1998), followed by a reduction of the matrixes' dimensionality by applying a projection over k predefined dimensions, similar to the least-squares method. Once the semantic space has been developed, the semantic distance between concepts or textual elements can be assessed through the calculation of their cosine similarity.¹

Topic relatedness through latent Dirichlet allocation

Similar to LSA, LDA is an NLP technique that provides information about the semantic content of text. This technique generates topic models based on an inference mechanism of underlying topic structures through a generative probabilistic process (Blei et al., 2003). On the basis of the assumption that documents consist of multiple topics, each document in a corpus is considered to consist of a random mixture of topics that occur throughout the entire corpus. A topic is a Dirichlet distribution (Kotz, Balakrishnan, & Johnson, 2000) over the space of thematically related terms that have similar probabilities of occurrences. Although each topic from the model contains all words with a corresponding probability, a remarkable demarcation can be observed between salient versus dominant concepts after the inference phase. As such, LDA topics reflect sets of concepts that co-occur more frequently (Blei & Lafferty, 2009). Despite the fact that LDA models rely on a few latent variables, exact inference is generally intractable (Heinrich, 2008). Therefore, approximate inference algorithms are used in practice, out of which Gibbs sampling (Griffiths, 2002) seems to be the most appropriate and

¹ Multiple optimizations can be considered to increase the reliability of the semantic vector-space representation, including the size of the training text corpora and the number of k dimensions after projection. The minimum size of the term-document matrix should be at least 20,000 terms with 20,000 passages (Landauer & Dumais, 2008). The optimal range for the number of dimensions k is 300 ± 50 , (Berry, Drmac, & Jessup, 1999; Jessup & Martin, 2001; Landauer, McNamara, Dennis, & Kintsch, 2007; Lemaire, 2009; Lizza & Sartoretto, 2001). Term frequency-inverse document frequency (TF-Idf; Manning & Schütze, 1999) and log-entropy (Landauer et al., 1998) are frequently used optimizations. Normalization of the word occurrences also improves performance. Stemming applied on all words reduces the overall performance, because each inflected form can express different perceptions and is related to different concepts, as discussed by Lemaire (2009) and Wiemer-Hastings and Zipitria (2001). Part-of-speech tagging is an additional consideration, but is of debatable value (Rishel, Perkins, Yenduri, & Zand, 2006; Wiemer-Hastings & Zipitria, 2001).

frequently used alternative. Because KL divergence (Kullback & Leibler, 1951) is not symmetric, therefore an improper distance measure, the inverse of the Jensen–Shannon dissimilarity (Cha, 2007; Manning & Schütze, 1999) can be used as a symmetrically smoothed alternative for expressing semantic similarity between textual fragments.²

Cohesion scoring mechanism for determining the importance of contributions

To obtain a quantitative analysis of participation, the importance of each contribution in relation to the overall conversation must be assessed by making use of the previously defined cohesion graph through CNA. Our aim is to assign a score to each utterance that reflects its overall coverage of the topic as well as the strength of relatedness between utterances in terms of its cohesion. This enables us to model the general trend of the overall conversation on the basis of the uttered concepts.

Therefore, the measured “impact” of each utterance is based on the underlying concepts’ relevance and the existing cohesive links to other contributions. Utterance scoring depends directly on the relevance of contained words. Thus, the score of each contribution is computed as the sum of the constituents’ relevance. As was presented in previous studies (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015b), two factors have been considered for evaluating each word’s relevance in relation to its corresponding textual fragment that can be either a sentence, an utterance, or the entire conversation. First, *statistical presence* is reflected by the normalized term frequency of the word within the textual fragment. Second, *semantic relatedness* accounts for the semantic similarity between the individual word and the entire textual fragment on the basis of the previously defined cohesion measures. After aggregating the two factors, key concepts or keywords for the conversation are automatically extracted as words that have the highest overall relevance.

Moving beyond individual words to utterance scoring, our model initially assigns an individual score for each contribution equal to the normalized term frequency of each constituent word, multiplied by its previously determined relevance (Dascalu, 2014). In other words, we measure the extent to which each utterance conveys the principal concepts of the overall conversation, as an estimation of on-topic relevance. Individual scores are subsequently augmented through cohesion links from CNA to other inter-linked contributions by

using cohesion values as weights. In other words, cohesive links from the cohesion graph are used to increase each utterance’s local importance score with the cumulative effect of other related contributions’ scores, multiplied by the corresponding semantic similarity values. To some extent, this process resembles eigenvector centrality in which the importance of each contribution is influenced by the strength of the cohesive links to other related contributions.

Overall, the assigned scores can be perceived as the importance of each textual element within the discourse, reflected as a mixture of both topic coverage and semantic relatedness to other textual elements. This mechanism can be easily extended through an extractive summarization algorithm that presents only the most important contributions from the conversation based on a threshold imposed by the user. This additional functionality was found to be useful by the tutors employing the functionalities of our earlier version of our system (Rebedea et al., 2010), but was not subject to formal validation.

The CNA sociogram

The CNA *sociogram* reflects the interaction between participants through cohesive links and, consequently, is an important data structure from which participation is assessed. The *sociogram* captures actor-actor ties and represents a collided view of the multithreaded polyphonic structure. Starting from the previously built CNA cohesion graph, we sum all link scores (i.e., individual contributions scores multiplied by the corresponding semantic similarity values) from the entire conversation between two speakers; the latter cumulative scores reflect the impact of the interchanged utterances between speakers (Dascalu, Chioasca, & Trausan-Matu, 2008). Instead of counting the exchanged utterances between participants, which can be considered the baseline in modeling actor-actor ties, our sociogram uses both the cohesion between the utterances, as well as their previously defined importance scores, to take into account the quality of the dialogue.

Starting from the sociogram, specific SNA metrics can be applied on the directed graph in order to measure centrality or participation. First, in-degree and out-degree centralities are computed as the sum of cohesive links to and from other participants. Although out-degree reflects each member’s active participation within the community, in-degree can be perceived as a form of popularity or prominence. Second, between-ness centrality (Bastian, Heymann, & Jacomy, 2009) reflects the status of central nodes that, if eliminated, would highly reduce or eliminate communication among other participants. In other words, those participants act as bridges for the information exchange between the members of the community. Third, closeness centrality (Sabidussi, 1966) represents the inverse distance to all other nodes in the graph;

² Similarly to LSA in which the number of dimensions k is pre-imposed, LDA has an imposed number of k topics, usually set to 100 as suggested by Blei et al. (2003). Teh, Jordan, Beal, and Blei (2006) have introduced hierarchical Dirichlet process (HDP), a nonparametric Bayesian approach also based on Dirichlet distributions to cluster grouped data. HDP is a generalization of LDA in which the number of topics is unbounded and inferred from the training text corpora (Teh et al., 2006), thus enabling groups to share statistical strength between clusters.

therefore, a higher value reflects a participant's stronger connection to other nodes.

However, the sociogram for CSCL chats has specific traits in contrast to other CSCL technologies due to the small number of participants (typically three to five students) and the large number of exchanged utterances. In this particular case, because we are dealing in most cases with a complete graph for chat conversations, between-ness scores for all nodes are 0. Centrality is also not a significant discriminant, because there was direct communication between all members. To further sustain this claim, Newman (2010) argues that closeness centrality is less useful than degree centrality from a mathematical point of view, because of its smaller range. However, the context changes dramatically for larger discussion groups, for which the usage of SNA metrics provides valuable insights in terms of participation, as we present in detail in the following section.

As an optimization of prior studies (Dascalu, Trausan-Matu, & Dessus, 2014; Trausan-Matu et al., 2014), adjacent utterances from the same participant within a limited timeframe (experimentally set to 1 min) were merged into a single contribution. In general, the tendency in chats is to separate the discourse into smaller textual units by demarcating each point as a new contribution; however, in most cases, adjacent utterances within a short timeframe that have the same speaker create a cohesive context. Moreover, this merging step is beneficial for subsequently applying the semantic models, more specifically LSA and LDA. Because these models are based on the bag-of-words approach and are adequate to be applied on larger textual elements for which this precondition is met, this optimization helps to create a more cohesive and dense representation of discourse. For the targeted chat experiments, this merge step reduced the global number of utterances by up to 20%. Explicit links manually

created by participants were transposed into the cohesion graph by adding links between the unified contributions.

In addition to the sociogram, the *evolution chart* in Fig. 3 is descriptive in terms of observing interaction patterns. The chart is based on the cumulative utterance scores, similar to the visualizations provided by Polyphony (Trausan-Matu, Rebedea, Dragan, & Alexandru, 2007) and A.S.A.P. (Dascalu et al., 2008). At each step of the conversation, the cumulated score for the speaker is increased with the importance score of the uttered contribution, thus modeling each member's participation up to a given moment. For example, zones with high slopes are indicative of monologues from a single participant, with diminished involvement of others. Figure 3 provides an example of a particular chat, namely the eighth conversation selected for our validation study, comprising four participants. In this example, all of the utterances from the conversation transcript with the identifiers from 220 up to 235 pertain solely to Participant 3; from 242 to 261, only two utterances do not belong to Participant 4, whereas Participant 1 completely dominated the discussion from 288 up to 300. Therefore, the generated graph clearly highlights zones with differential involvement of participants within the ongoing conversation—that is, monologue of one participant and the stagnation of all other members' evolution lines. A more suitable configuration for participation within the given instructional setting in which users should have collaborated with each other would contain comparable growths of multiple participants. This translates into a more equitable involvement of multiple speakers, similar to the situations presented in both chat excerpts from Tables 1 and 2.

Overall, on the basis of the previous analysis, four quantitative indices for participation emerge: (a) *cumulative utterance scores* per participant (i.e., the sum of individual contribution scores that were uttered by a certain participant), as

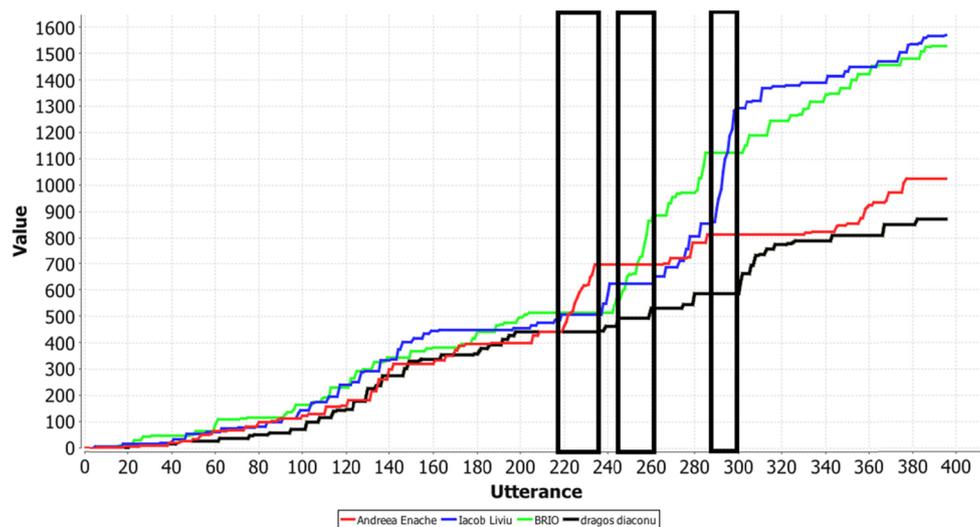


Fig. 3 Evolution chart highlighting monologues of certain participants

well as (b) *in-degree*, and (c) *out-degree* SNA metrics (i.e., sum of scores corresponding to inbound and outbound edges for a given node) computed from the sociogram. This can also be perceived as an extension on CNA in terms of modeling the interaction between different participants in a polyphonic manner, through cohesion.

Integration of multiple CNA graphs

Starting from the analysis of a single conversation at a time, our model can be further extended to facilitate the evaluation of online communities by generalizing the assessment of isolated threads to an aggregation facility of multiple discussion threads. Similar to the process introduced by Suthers and Rosen (2011) of constructing a global network based on traces from fragmented logs, we enable the evaluation of participation at a macroscopic level, not only at the level of individual discussions. The discrepancy between a local view and a global one has multiple implications, and specific technical aspects need to be taken into consideration when merging multiple discussion threads. From a technical perspective, the shift requires the aggregation of individual utterance scores from each conversation CNA cohesion graph and building a global sociogram of all unique participants on which SNA metrics can be applied. Overall, the exploration of different user

distributions, goals, and configurations is intended to offer a broader perspective on how participation evolves and to employ our CNA model in different educational scenarios.

For example, the CNA sociogram depicted in Fig. 4 was generated on the basis of 444 forum discussion threads summing 3,685 contributions that spanned August 2010 to June 2012 (Nistor et al., 2014). The conversation threads included 179 participants (20 full-time faculty employees and 159 part-time faculty members), all of them holding a doctoral degree. A clear demarcation can be observed between different types of users: For example, Member 16 is by far the most actively involved member (415 contributions, vs. the second and third most active members, Member 29 [255] and Member 25 [229]), whereas members with lower participation (e.g., Member 55, 143 contributions) tend to have a more peripheral position. Of course, members with few contributions are close to the outer bounds of the community, since they only have a limited number of connections to other members in the global CNA sociogram.

In contrast, as is reflected in the snapshots from Fig. 5, online communities develop differently, depending of the environment, their central members, and the covered topics. For example, Figure 5a depicts the CNA sociogram for a dense community with approximately 450 overall users (throughout its entire lifespan) and 290 contributions in 1 year, in contrast to a user-centered

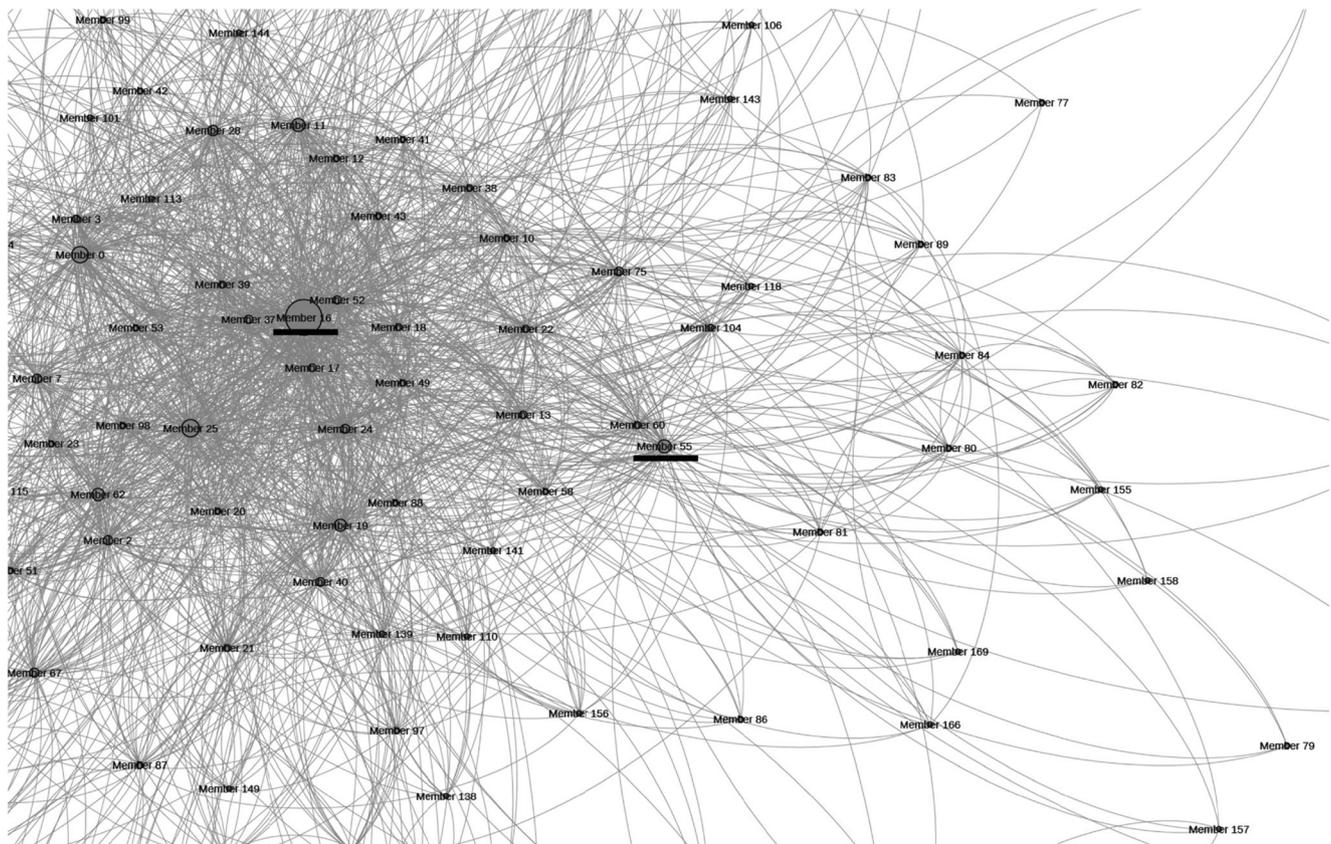


Fig. 4 Partial view of a CNA sociogram corresponding to an academic forum

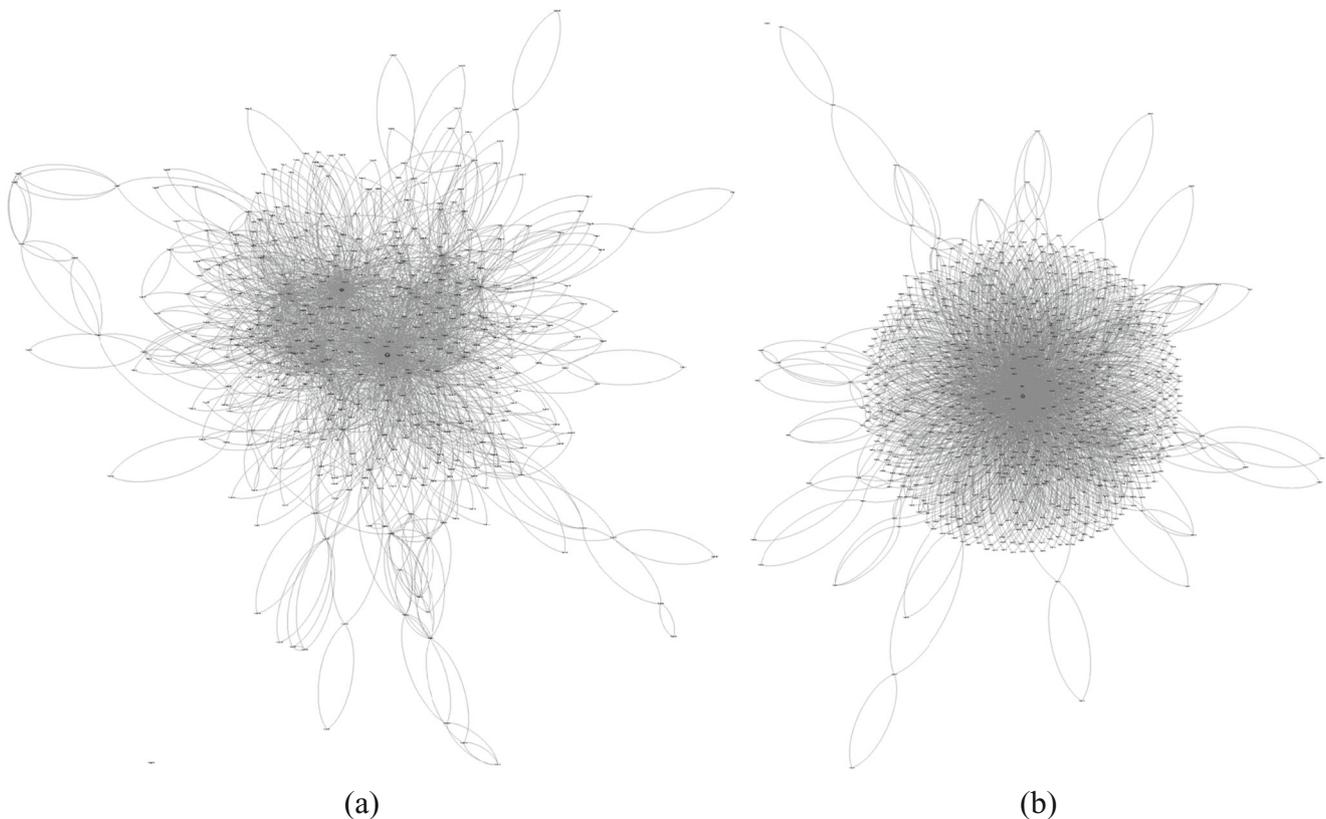


Fig. 5 CNA sociograms for (a) a dense community and (b) a user-centered community

community with approximately 750 members and 550 contributions in 1 year (see Fig. 5b). A major difference can be observed between the topologies of the sociograms corresponding to the two communities. The first graph consists of multiple discussion threads governed by multiple participants, in which the core of the community contains active and central members (Nistor, Dascalu, & Trausan-Matu, 2016; Nistor et al., 2015), whereas the second presents a radial view centered on one individual—the blog owner.

The concepts of keywords and main group topics, as well as central members, are common to all configurations; still, their influences are very different for each type of analyzed group. Thus, the lifecycle of a member within chats is relatively short (measured in hours), whereas the span increases in forums to each thread's lifetime. Finally, online community members “live” from their first post up until the community's lifetime ends. Thus, we can say that a member's *centrality* can be better monitored in long-term discussion threads (e.g., forums, online communities), whereas topic specificity and coverage is modeled better in small chats or discussion threads.

Validation study

The proposed CNA model enables an in-depth, cohesion-centered evaluation of learners' active engagement within CSCL environments. Participation plays a key role for instructors

in CSCL learning scenarios (Lehtinen, 2003) as it represents the building block of collaboration which, in turn, involves the mutual engagement of participants in their effort to jointly solve an endeavor (i.e., collaborative problem-solving tasks; Roschelle & Teasley, 1995; Stahl, 2006, 2009). Moreover, of particular interest is the free-rider effect commonly encountered in CSCL environment in which a member allows other people to do their work (Dillenbourg, 2002; Salomon & Globerson, 1989), thus creating a discrepancy in terms of participation. Our CNA model provides the mechanisms to timely assess the distribution of participation among chat members, therefore identifying major differences that are indicative of the free-rider effect.

Corpus selection

The corpus used for the validation of our CNA model comprises ten chat conversations selected from a larger corpus of over 100 undergraduate student chats that took place in an academic environment. In the first part of these conversations, four to five Computer Science undergraduate students debated on the benefits and disadvantages of specific CSCL technologies (e.g., chat, blog, wiki, forum, or Google Wave). Each chat had an equitable gender distribution and the participants previously knew each other by attending the same course. Each student was the advocate of a given technology by trying to convince the other participants of its advantages in contrast

to the alternatives. In the second part, all participants were asked to jointly propose an integrated platform as a viable alternative to be used by a company, which in turn encompassed most of the previously presented advantages.

Our aim was to consider different types of conversations in terms of (a) the length of the conversation, expressed in number of utterances, (b) the frequency of utterances (i.e., contributions per minute), and (c) specific participation and interaction patterns (e.g., domination in turn of each participant, disproportionate involvement versus equitable participation of all members in the ongoing conversation). The specific characteristics of each of the ten selected conversations are presented in Table 3. The particular cases, derived from the previous criteria and matched by our automated CNA model, are discussed in detail in the Results section.

To grasp the specific language used within the conversations, LSA and LDA semantic models were trained on a corpus containing both the TASA corpus (Touchstone Applied Science Associates, Inc., <http://lsa.colorado.edu/spaces.html>), for general background knowledge, and a collection of more than 500 CSCL-related scientific articles. Paragraphs with fewer than 20 content words were disregarded.

Human judgments

Human judgments of the target corpus were used to validate the automated indices of the CNA participation model. To accomplish the time-consuming process of manual annotation, the evaluation process was assigned to four raters (two undergraduate and two graduate students in Computer Science) who were asked to assess the participation of each speaker on a Likert scale of 1 to 10. Participation was rated along two dimensions and speakers received two scores: one denoting their active involvement throughout the entire conversation and another reflecting their interaction with the other participants. To achieve this, each coder received an Excel spreadsheet with ten sheets (one for each chat transcript) in which they were asked to fill in the individual participation scores for the speakers. On the basis of all 47 participant ratings, the average intraclass correlation (ICC) was .661 and Cronbach's alpha was .749 for active involvement, and the average ICC was .553 and Cronbach's alpha was .620 when quantifying interaction with other participants. As expected, reliability was lower for interaction, as this task was a more subjective and error-prone assessment considering the length of the conversations.

Results

Nonparametric correlations (Spearman's Rho) were calculated between either the overall number of contributions (the baseline of this analysis) or the automatically computed CNA indices and the

Table 3 Descriptive statistics for the conversation corpus

Conversation	Utterances	Participants	Time Span	Utterance Frequency/Minute	Observations
Chat 1	339	5	1 h 50 min	3.08	Presence of two low engaged students
Chat 2	283	5	1 h 15 min	3.77	Equitable involvement
Chat 3	405	5	2 h	3.38	Focus on one technology, dominance of one active student (>100 contribution), major disequilibrium between participants
Chat 4	251	5	1 h 35 min	2.64	Presence of one low involved student
Chat 5	416	5	1 h 35 min	4.38	Off-balance induced by the presence of one low involved student and one active student (>100 contribution)
Chat 6	378	5	1 h 30 min	4.20	Two active, two moderately active, and one low engaged student
Chat 7	270	5	1 h 40 min	2.70	Relatively low engaged students overall
Chat 8	389	4	1 h 50 min	3.54	Monologues, in turn, by certain participants
Chat 9	190	4	45 min	4.22	Shortest conversation with the lowest overall participation scores
Chat 10	297	4	1 h 2 5min	3.49	Actively involved students with an equitable participation
Avg (Stdev)	321.8 (75.24)	4.70 (0.48)	1 h 33 min (21 min)	3.54 (0.62)	

average rater scores for all conversations, since our intent was to create a general predictive model for all conversations. Spearman's Rho correlations were used instead of Pearson correlations because we were focused on the rankings of the participants throughout all the conversations; see Table 4 for these results.

The chat results demonstrate the reliability and adequacy of the proposed quantitative indices in assessing participation, as well as the complementarity of the implemented indices. Although involvement scores are highly correlated with the number of contributions, thus denoting a strong quantitative limitation of the manual assessment, our CNA model proves to be more informative in terms of the interaction between participants. However, we are inherently limited by the quantitative dimension of our analysis—the more users speak (without necessarily being on topic), the greater their impact and the higher their chance of shifting the central view of the whole conversation.

Two stepwise regression analyses were performed in order to determine the degree to which the automated indices predicted the human ratings of participation. These regressions yielded significant models: $F(2, 44) = 16.636, p < .001, r = .656, R^2 = .431$, for involvement, and $F(1, 45) = 34.979, p < .001, r = .661, R^2 = .437$, for interaction. Two variables were significant predictors in the regression analysis for involvement, accounting for 43% of the variance in the manual annotations: *cumulative utterance scores* [$\beta = 2.464, t(2, 44) = 5.007, p < .001$] and *out-degree* [$\beta = -2.071, t(2, 44) = -4.210, p < .001$]. One variable was a significant predictor in the regression analysis for interaction, accounting for 44% of the variance in the manual annotations: *cumulative utterance scores* [$\beta = .661, t(1, 45) = 5.914, p < .001$]

In contrast, when calculated in relation to the baseline (i.e., number of contributions), the regression analyses also yielded significant models: $F(1, 45) = 34.377, p < .001, r = .658, R^2 = .433$, for involvement, and $F(1, 45) = 22.528, p < .001, r = .578, R^2 = .334$, for interaction. Although the amount of variance accounted for is similar to that of our model in terms of involvement, CNA clearly explains considerably more variance in terms of the interaction among participants.

Discussion

We have introduced a computational model of participation in CSCL conversations based on CNA. Cohesion describes the

information transfer between participants and measures the topic relatedness between utterances, whereas participation is automatically assessed as members' active involvement in cohesive contexts. This model can be used to conduct just-in-time assessments, providing the potential to intervene and facilitate equitable involvement by participants in CSCL and team-based learning environments. On the basis of the performed comparative analyses of different CSCL environments, topics and utterance scores tend to have a greater impact on local analyses (i.e., single conversations, chats), whereas centrality measures gain a greater importance in the evaluation of global sociograms. The latter are obtained after integrating multiple CNA cohesion graphs and the importance of each member within the global sociograms is better reflected through their centrality score. Moreover, specific triggers can be added in order to encourage interactions among participants if nonoptimal patterns (e.g., the free-rider effect) are observed, thus converging toward more desirable outcomes of equitable involvement of members within CSCL conversations (Dascalu, 2014; Hoadley, 2002). In addition, when generalized and applied to informal online knowledge communities, our CNA model can be used to classify and characterize different types of users (e.g., central, active, peripheral community members; Nistor et al., 2016; Nistor et al., 2015).

Comparison to other CSCL discourse models

Besides highlighting the extensions and wide applicability of our CNA model, we must also consider a comparison to other CSCL models, namely the contingency graph (Medina & Suthers, 2009; Suthers, 2015; Suthers & Desiato, 2012), transactivity (Joshi & Rosé, 2007; Rosé et al., 2008), epistemic network analysis (ENA; Shaffer et al., 2009) and the model of constructing networks of action-relevant episodes (CN-ARE; Barab, Hay, & Yamagata-Lynch, 2001). First, the *contingency graph* is used as the basis for representing transcriptions and highlights contingencies between events. The contingency graph relies on: (a) generic events that can be traced to the interaction with the CSCL environment, including the creation, manipulation and perception of media inscriptions, and (b) contingency relationships when one or more events enable a subsequent event. Our CNA model uses textual contributions and follows the multithreaded

Table 4 Spearman Rho correlations between the indices and mean rater participation (for both involvement and interaction scores) for all conversations combined

Participation Dimension	Contributions	Cumulative Utterance Scores	In-Degree	Out-Degree
Involvement	.671**	.498**	.285	.373**
Interaction	.618**	.683**	.631**	.652**

* $p < .05$. ** $p < .01$

polyphonic structure, reflecting the overall discourse cohesion derived from the interactions between participants. As an analogy, CNA reflects in an automated manner the *temporal proximity* between events that create media inscriptions and provides a deeper understanding of lexical contingencies (that typically only analyze the number of overlapping stems) by relying on multiple semantic models.

Second, transactivity (Joshi & Rosé, 2007) can be perceived as a complementary view to our CNA approach centered on the information flow between participants through cohesion. In this view, transacts highlight the relationships between competing positions of different speakers, similar to dialogue acts (Stolcke et al., 2000), but at a different semantic granularity. Therefore, transacts can become a potential extension of our CNA model in which cohesion, corroborated with opinion mining and sentiment analysis, could be used to evaluate the convergence or divergence of participants' points of view.

Third, Collier, Ruis, and Shaffer (2016) have extended their model of ENA to evaluate connections within discourse. ENA is a method that identifies and evaluates connections among elements in coded data, both visually and through statistics, by representing links in dynamic networks. ENA aims at evaluating learning in CSCL environments by also considering local relationships between concepts and patterns in discourse, within a given domain. Similar to ENA, CN-ARE (Barab et al., 2001) also provides a broader methodological context that facilitates the identification of relevant data from a complex, evolving environment, followed by its organization into a web of action with its corresponding evolving trajectory. In contrast to the latter two models, our CNA model is grounded in automated text analysis and semantic models that can facilitate a deeper understanding of discourse and the cohesive links among text segments. Thus, our model could be enriched with other artifacts to provide a more generalized perspective, similar to ENA.

Limitations and extensibility

Although it is successful, we must highlight certain limitations of our model. In particular, this demonstration of the model was employed within a specific educational context in which participants share, continue, debate, or argue certain topics or key concepts of the conversation. This educational context creates the premises for building cohesive conversation threads in which the importance of each contribution is correspondingly augmented via CNA. In addition, all computational perspectives are inevitably limited when analyzing the dialogical nature of discourse: “it is indeed impossible to be “completely dialogical”, if one wants to be systematic and contribute to a cumulative scientific endeavor” (Linell, 2009, p. 383). Certain discourse segments might be dominated by a participant whose overall score will be augmented in an

artificial manner. In contrast, active collaboration among peers on off-topic concepts, generating discussion threads of low relevance, will be detrimental to the overall evaluation framework as the automatically extracted keywords would have also shifted.

Overall, in this article, we have transcended traditional SNA metrics applied on the exchanged utterances, to CNA, which emphasizes the importance of semantics in discourse analysis. Therefore, CNA takes network analysis further by explicitly considering semantic cohesion while modeling interactions between participants. In addition, our experiment highlights the benefits of using specific NLP tools in social interaction domains and for modeling the underlying discourse structure. Overall, the presented model should not be perceived as a rigid structure, but as an adaptable one that evolves on the basis of the cohesion to other participants' utterances. Additionally, CNA is a highly flexible model that can be used to model a wide range of CSCL scenarios, covering various environments in which participants are encouraged to collaborate and exchange ideas.

Conclusions and future research directions

The model and validation presented in this study revealed that a cohesion-based discourse structure can be used to perform an in-depth analysis of participation within multiple educational contexts: chats, forums, or online knowledge-building communities of practice. Furthermore, the CNA model has demonstrably strong potential to be successfully extended to other configurations such as forums and other online communities. Besides introducing a fully automated assessment method relying on advanced NLP techniques, the strength of our CNA approach consists of modeling polyphonic structures in the interaction among participants through text cohesion.

The aim to assess participation from a quantitative point of view cannot be obtained by using a single index. Our combined CNA model provides reliable estimations of participation, as compared to human, manual assessments, while specific indices have the potential to reflect individual traits that evolve in time. In addition, the discourse traits reflected by textual cohesion represent a building block for measuring participation at a macroscopic level.

As a further extension of this model, we envision the introduction of additional participation indices based on the voices' coverage for each conversational participant to deepen the dialogical framing of our analysis. To further refine the prediction of active involvement derived from dialogism, we will explore implicit links derived from additional semantic models (e.g., word2vec; Mikolov, Chen, Corrado, & Dean, 2013), as well as interaction patterns derived from speech acts (Searle, 1969). Moreover, we will consider quantifying the difference between types of networks using standard metrics,

namely (a) mean geodesic distance (for individual conversations), (b) clustering coefficient or transitivity, and (c) degree distributions (Barabási, 2016; Newman, 2010). Our long-term objective consists of broadening our perspective to an even more automated environment where students and tutors alike can self-assess the quality of collaboration and performance using our integrated framework, ReaderBench.

Author note We thank the students of University “Politehnica” of Bucharest who participated in our experiments and to Lucia Larise Stavarache for her support in processing the chat conversations. This research was partially supported by the FP7 208-212578 LTLL project, by the 644187 RAGE H2020-ICT-2014 project, as well as by the NSF 1417997 and 1418378, and the Office of Naval Research (ONR N000141410343) grants to Arizona State University.

References

- Amsteth, H. C., & Ludvigsen, S. (2006). Approaching institutional contexts: Systemic versus dialogic research in CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 167–185.
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin, TX: University of Texas Press.
- Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics* (C. Emerson, Trans. C. Emerson Ed.). Minneapolis, MN: University of Minnesota Press.
- Barab, S. A., Hay, K. E., & Yamagata-Lynch, L. C. (2001). Constructing Networks of Action-Relevant Episodes: An In Situ Research Methodology. *Journal of the Learning Sciences, 10*, 63–112. doi:10.1207/S15327809JLS10-1-2_5
- Barabási, A. L. (2016). *Network Science*. Cambridge, UK: Cambridge University Press.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media* (pp. 361–362). San Jose, CA: AAAI Press.
- Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.
- Berry, M. W., Drmac, Z., & Jessup, E. R. (1999). Matrices, vector spaces, and information retrieval. *SIAM Review, 41*, 335–362.
- Blei, D. M., & Lafferty, J. (2009). Topic Models. In A. Srivastava & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications* (pp. 71–93). London, UK: Chapman & Hall/CRC.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.
- Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics, 32*, 13–47.
- Cassirer, E. (1953). *The philosophy of symbolic forms* (Vol. 1). New Haven, CT: Yale University Press.
- Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences, 1*, 300–307.
- Collier, W., Ruis, A., & Shaffer, D. W. (2016). Local versus global connection making in discourse. In 426–433 (Ed.), *12th International Conference on Learning Sciences (ICLS 2016)*. Singapore: International Society of the Learning Sciences (ISLS).
- Cress, U. (2013). Mass collaboration and learning. In R. Luckin, S. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 416–424). New York, NY: Routledge.
- Dascalu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating* (Studies in Computational Intelligence, Vol. 534). Berlin, Germany: Springer.
- Dascalu, M., Chioasca, E. V., & Trausan-Matu, S. (2008). ASAP—An Advanced System for Assessing Chat Participants. In D. Dochev, M. Pistore, & P. Traverso (Eds.), *13th International Conference on Artificial Intelligence: Methodology, systems, and applications (AIMSA 2008)* (pp. 58–68). Berlin, Germany: Springer.
- Dascalu, M., Trausan-Matu, S., & Dessus, P. (2013). Cohesion-based analysis of CSCL conversations: Holistic and individual perspectives. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *10th International Conference on Computer-Supported Collaborative Learning (CSCL 2013)* (pp. 145–152). Madison, USA: ISLS.
- Dascalu, M., Trausan-Matu, S., & Dessus, P. (2014). Validating the automated assessment of participation and of collaboration in chat conversations. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *12th International Conference on Intelligent Tutoring Systems (ITS 2014)* (pp. 230–235). Honolulu, USA: Springer.
- Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench—Automated Evaluation of Collaboration on the basis of Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning, 10*, 395–423. doi:10.1007/s11412-015-9226-y
- Dascalu, M., Stavarache, L. L., Dessus, P., Trausan-Matu, S., McNamara, D. S., & Bianco, M. (2015). ReaderBench: The learning companion. In *17th International Conference on Artificial Intelligence in Education (AIED 2015)* (pp. 915–916). Madrid, Spain: Springer.
- Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D. S. (2015a). Dialogism: A framework for CSCL and a signature of collaboration. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *11th International Conference on Computer-Supported Collaborative Learning (CSCL 2015)* (pp. 86–93). Gothenburg, Sweden: ISLS.
- Dascalu, M., Stavarache, L. L., Trausan-Matu, S., Dessus, P., Bianco, M., & McNamara, D. S. (2015b). ReaderBench: An integrated tool supporting both individual and collaborative learning. In *5th International Learning Analytics & Knowledge Conference (LAK'15)* (pp. 436–437). New York, NY: ACM.
- Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D. S. (2015b). Discourse cohesion: A signature of collaboration. In *5th International Learning Analytics & Knowledge Conference (LAK'15)* (pp. 350–354). Poughkeepsie, NY: ACM.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Harshman, R., Landauer, T. K., Lochbaum, K., & Streeter, L. (1989). USA Patent No. 4,839,853. 4,839,853: USPTO.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science, 41*, 391–407.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Heerlen, The Netherlands: Open Universiteit Nederland.
- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology, 38*, 188–230.
- Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik, 14*, 403–420.
- Griffiths, T. (2002). *Gibbs sampling in the generative model of latent Dirichlet allocation*. Stanford, CA: Stanford University.

- Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 203–225.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion In English*. London, UK: Longman.
- Heinrich, G. (2008). *Parameter estimation for text analysis*. Leipzig, Germany: vsonix GmbH + University of Leipzig.
- Hoadley, C. P. (2002). *Creating context: Design-based research in creating and understanding CSCL*. Paper presented at the International Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community, Boulder, Colorado.
- Hobbs, J. R. (1978). *Why is discourse coherent?* Menlo Park, California: SRI International.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67–90.
- Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Center for the Study of Language and Information: Stanford University.
- Hobbs, J. R. (1990). Topic drift. In B. Dorval (Ed.), *Conversational organization and its development* (pp. 3–22). Norwood, NJ: Ablex.
- Holmer, T., Kienle, A., & Wessner, M. (2006). Explicit Referencing in Learning Chats: Needs and Acceptance. In W. Nejdl & K. Tochtermann (Eds.), *Innovative approaches for learning and knowledge sharing: First European Conference on Technology Enhanced Learning, EC-TEL 2006* (pp. 170–184). Crete, Greece: Springer.
- Jessup, E. R., & Martin, J. H. (2001). Taking a new look at the Latent Semantic Analysis approach to information retrieval. In M. W. Berry (Ed.), *Computational information retrieval* (pp. 121–144). Philadelphia, PA: SIAM.
- Joshi, M., & Rosé, C. P. (2007). *Using transactivity in conversation summarization in educational dialog*. Paper presented at the SLATE Workshop on Speech and Language Technology in Education, Farmington, Pennsylvania, USA.
- Jurafsky, D., & Martin, J. H. (2009). *An introduction to Natural Language Processing. Computational linguistics, and speech recognition* (2nd ed.). London, UK: Pearson Prentice Hall.
- Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In C. M. Hoadley & J. Roschelle (Eds.), *International Conference on Computer Support for Collaborative Learning (CSCL'99)* (pp. 308–313). Palo Alto: ISLS.
- Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). Dirichlet and inverted Dirichlet distributions. In *Continuous multivariate distributions: Vol. 1: Models and applications* (2nd ed., pp. 485–527). New York, NY: Wiley.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–240. doi:10.1037/0033-295X.104.2.211
- Landauer, T. K., & Dumais, S. (2008). Latent semantic analysis. *Scholarpedia*, 3, 4356.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. doi:10.1080/01638539809545028
- Landauer, T. K., Laham, D., & Foltz, P. W. (1998). Learning human-like knowledge by singular value decomposition: A progress report. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), *Advances in Neural Information Processing Systems* (Vol. 10, pp. 45–51). Cambridge, MA: MIT Press.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- Lehtinen, E. (2003). Computer-supported collaborative learning: An approach to powerful learning environments. In E. De Corte, L. Verschaffel, N. Entwistle, & J. Van Merriëboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 35–54). Amsterdam, The Netherlands: Elsevier.
- Lemaire, B. (2009). Limites de la lemmatisation pour l'extraction de significations. In *9es Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2009)* (pp. 725–732). Lyon, France: Presses Universitaires de Lyon.
- Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: Information Age.
- Lizza, M., & Sartoretto, F. (2001). A comparative analysis of LSI strategies. In M. W. Berry (Ed.), *Computational information retrieval* (pp. 171–181). Philadelphia, PA: SIAM.
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. Marina del Rey, CA: Information Sciences Institute.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Marková, I., Linell, P., Grossen, M., & Salazar Orvig, A. (2007). *Dialogue in focus groups: Exploring socially shared knowledge*. London, UK: Equinox.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47, 292–330.
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.
- Medina, R., & Suthers, D. (2009). Using a contingency graph to discover representational practices in an online collaborative environment. *Research and Practice in Technology Enhanced Learning*, 4, 281–305.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representation in vector space. In *Workshop at ICLR*. Scottsdale, AZ.
- Newman, M. E. J. (2010). *Networks: An introduction* (1st ed.). Oxford, UK: Oxford University Press.
- Nistor, N., Baltes, B., Dascalu, M., Mihaila, D., Smeaton, G., & Trausan-Matu, S. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior*, 34, 339–344. doi:10.1016/j.chb.2013.10.051
- Nistor, N., Trausan-Matu, S., Dascalu, M., Duttweiler, H., Chiru, C., Baltes, B., & Smeaton, G. (2015). Finding student-centered open learning environments on the internet: Automated dialogue assessment in academic virtual communities of practice. *Computers in Human Behavior*, 47, 119–127. doi:10.1016/j.chb.2014.07.029
- Nistor, N., Dascalu, M., & Trausan-Matu, S. (2016). Newcomer integration in online knowledge communities: Exploring the role of dialogic textual complexity. In *12th Int. Conf. on Learning Sciences (ICLS 2016)* (pp. 914–917). Singapore: International Society of the Learning Sciences (ISLS).
- Rebedea, T. (2012). *Computer-based support and feedback for collaborative chat conversations and discussion forums* (Doctoral dissertation). University Politehnica of Bucharest, Bucharest, Romania.
- Rebedea, T., Dascalu, M., Trausan-Matu, S., Banica, D., Gartner, A., Chiru, C. G., & Mihaila, D. (2010). Overview and preliminary results of using PolyCAFe for collaboration analysis and feedback generation. In M. Wolpers, P. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Sustaining TEL: From innovation to learning and practice: 5th European Conference on Technology Enhanced Learning (EC-TEL 2010)* (pp. 420–425). Barcelona, Spain: Springer.
- Rishel, T., Perkins, A. L., Yenduri, S., & Zand, F. (2006). Augmentation of a term/document matrix with part-of-speech tags to improve accuracy of latent semantic analysis. In *5th WSEAS International Conference on Applied Computer Science* (pp. 573–578). Hangzhou, China.

- Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-Supported Collaborative Learning*. New York, NY: Springer.
- Rosé, C. P., 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, 237–271.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31, 581–603.
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, 13, 89–100.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97–118). New York, NY: Cambridge University Press.
- Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.
- Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E.,... Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *IJLM*, 1, 33–53.
- Stahl, G. (2006). *Group cognition. Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.
- Stahl, G. (2009). *Studying virtual math teams*. New York, NY: Springer.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.
- Stahl, G., Cress, U., Ludvigsen, S., & Law, N. (2014). Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning*, 9, 117.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, J., Bates, R., Jurafsky, D.,... Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26, 339–373.
- Suthers, D. (2015). From contingencies to network-level phenomena: Multilevel analysis of activity and actors in heterogeneous networked learning environments. In *5th International Learning Analytics & Knowledge Conference (LAK'15)* (pp. 368–377). Poughkeepsie, NY: ACM.
- Suthers, D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *45th Hawaii International Conference on System Sciences* (pp. 3368–3377). Piscataway, NJ: IEEE Press.
- Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In *1st International Learning Analytics & Knowledge Conference (LAK'11)* (pp. 64–74). New York, NY: ACM.
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101, 1566–1581.
- Trausan-Matu, S. (2010a). Automatic support for the analysis of online collaborative learning chat conversations. In P. M. Tsang, S. K. S. Cheung, V. S. K. Lee, & R. Huang (Eds.), *3rd International Conference on Hybrid Learning* (pp. 383–394). Berlin, Germany: Springer.
- Trausan-Matu, S. (2010b). The polyphonic model of hybrid and collaborative learning. In F. Wang, L. J. Fong, & R. C. Kwan (Eds.), *Handbook of research on hybrid learning models: Advanced tools, technologies, and applications* (pp. 466–486). Hershey, NY: Information Science.
- Trausan-Matu, S., & Rebedea, T. (2009). Polyphonic inter-animation of voices in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 451–473). New York, NY: Springer.
- Trausan-Matu, S., & Rebedea, T. (2010). A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In A. F. Gelbukh (Ed.), *11th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2010)* (pp. 354–363). New York, NY: Springer.
- Trausan-Matu, S., Stahl, G., & Zemel, A. (2005). *Polyphonic inter-animation in collaborative problem solving chats*. Philadelphia, PA: Drexel University.
- Trausan-Matu, S., Rebedea, T., Dragan, A., & Alexandru, C. (2007). Visualisation of learners' contributions in chat conversations. In J. Fong & F. L. Wang (Eds.), *Blended learning* (pp. 217–226). Singapore: Pearson/Prentice Hall.
- Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2007). Supporting polyphonic collaborative learning. *E-Service Journal*, 6, 58–74.
- Trausan-Matu, S., Rebedea, T., & Dascalu, M. (2010). Analysis of discourse in collaborative learning chat conversations with multiple participants. In D. Tufis & C. Forascu (Eds.), *Multilinguality and interoperability in language processing with emphasis on Romanian* (pp. 313–330). Bucharest, Romania: Editura Academiei.
- Trausan-Matu, S., Dascalu, M., & Dessus, P. (2012). Textual complexity and discourse structure in Computer-Supported Collaborative Learning. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *11th International Conference on Intelligent Tutoring Systems (ITS 2012)* (pp. 352–357). Chania, Greece: Springer.
- Trausan-Matu, S., Dascalu, M., & Rebedea, T. (2014). PolyCAFe—Automatic support for the polyphonic analysis of CSCL chats. *International Journal of Computer-Supported Collaborative Learning*, 9, 127–156. doi:10.1007/s11412-014-9190-y
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
- Wegerif, R. (2005). A dialogical understanding of the relationship between CSCL and teaching thinking skills. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Conference on Computer Supported Collaborative Learning 2005 (CSCL'05): The next 10 years!* (p. 7). Taipei, Taiwan: ISLS.
- Wertsch, J. (1998). *Mind as action*. Oxford, UK: Oxford University Press.
- Wiemer-Hastings, P., & Zipitria, I. (2001). Rules for syntax, vectors for semantics. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 1112–1117). Mahwah, NJ: Erlbaum.
- Wu, Z., & Palmer, M. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics, ACL '94* (pp. 133–138). New York, NY: ACL.