Connectivity for Knowledge Building: A Framework of Socio-Semantic Network Motif Analysis

Bodong Chen, University of Minnesota, chenbd@umn.edu

Abstract: Productive knowledge-building discourse requires students to work as a collective to solve shared problems. Students need to engage with each other socially while continually producing ideas, debugging understanding, and rising above multiple ideas to build high-level knowledge structures. To investigate complex socio-cognitive dynamics in knowledge-building discourse, this paper proposes to model discourse as a socio-semantic network (SSN) and then use *network motifs*—defined as recurring, significant subgraphs—to characterize the network and the discourse. A framework of SSN motif analysis is proposed according to knowledge-building discourse processes. To demonstrate the utility of the framework, we applied it to discourse data from two classes. Results revealed distinct motif profiles of two classes across two discourse phases. The framework of SSN motif analysis shows promise as a novel, theoretically informed approach to discourse analysis.

Keywords: network analysis, knowledge building, discourse analysis, learning analytics

Introduction

Interpersonal communication is essential for learning in various contexts. Since the rise of the socio-cognitive paradigm in the 1980s, the importance of social interaction is emphasized in many educational theories and practices. In computer-supported collaborative learning (CSCL), learners are asked to not only nurture productive interdependence through social configurations such as "jigsaw groups," but also to learn collaboratively by constructing a shared problem space, building on one another, and creating shared knowledge artifacts (Roschelle & Teasley, 1995; Rosé et al., 2017). In collaborative discourse, learners are engaged in discussing substantive content related to a domain; by leveraging the interpersonal/intersubjective space, learners are expected to make sense of new concepts and build shared knowledge beyond the reach of each individual.

Knowledge Building (KB) as a pedagogy aims to engage students in productive collaborative discourse that leads towards continual improvement of ideas within a knowledge community (Chen & Hong, 2016). Discourse—face-to-face and/or online—is central to KB; discourse is where ideas are put forward, inspected, discussed, and improved. To sustain knowledge-building discourse, a set of twelve principles are put forward to guide discursive practices (Scardamalia, 2002). For instance, *idea diversity* is one important KB principle that states that knowledge advancement relies on a diverse set of ideas in the community; *improvable ideas*, another KB principle, depends on idea diversity, as well as students' *epistemic agency* and *collective responsibility* over their community knowledge. *Embedded and transformative assessment* plays a role in supporting idea improvement in knowledge-building discourse and has inspired the development of analytic tools to support students' reflective assessment and "meta-discourse" (discourse about discourse) (Resendes et al., 2015; Yang et al., 2021). To foster conditions for productive knowledge-building discourse, researchers, educators, and designers have worked concernedly to design and test various pedagogical strategies, software features, and analytic tools (Chen & Hong, 2016). Knowledge Forum, the most widely used knowledge-building environment, and its analytic tools are designed and continually refined to support KB (Scardamalia, 2004).

To support embedded assessment in knowledge-building discourse, the KB community has a long-standing interest in developing analytic tools to capture meaningful indicators from discourse. Many analytic tools have been developed by the community in order to help monitor discussion behaviors, analyze social networks, and evaluate idea development (e.g., Oshima et al., 2012; Yang et al., 2021; Zhang et al., 2018). The Analytic Toolkit developed in the 1990s provides a series of extracted analytics about reading, writing, and collaboration in knowledge-building discourse (Burtis, 1998). Knowledge Forum is also embedded with a suite of analytic "applets" powered by lexical analysis, social network analysis, and so on (Teplovs et al., 2007). Knowledge Building Discourse Analyzer (KBDeX), a more recent innovation, transforms written discourse into bipartite networks of learners and words, enabling the evaluation of knowledge-building discourse based on network measures (Oshima et al., 2012).

These efforts are in line with a strong interest in applying computational methods to discourse data in the CSCL and learning analytics communities. Colleagues have been applying a range of methods to examine collaborative discourse, with a particular focus on the social and cognitive domains of learning. Social network analysis (SNA)—both network metrics and visualizations—are broadly used to investigate social interaction in collaboration (Dawson, 2008). Computational content analysis is also used to examine cognitive content in collaborative discourse (Kovanović et al., 2017). Interesting efforts are also made to bridge qualitative coding of discourse data and quantitative summarization of discourse patterns (Shaffer, 2017).

Network analysis is a salient theme in this body of work. Network analysis methods are widely used to examine the cognitive, social, and integrated domains of collaborative discourse. Questions concerning the social aspect of discourse lend themselves to SNA. Social network centrality measures (such as degree, closeness, and betweenness) are used to evaluate a student's social position in a class and then correlate with students' sense of community (Dawson, 2008). Network methods are also used to analyze and represent cognitive content. For example, content entities such as words and artifacts can be modeled as networks. Prior work has leveraged different types of word relations, such as co-occurrence, word sequence, machine-learned similarity (Paranyushkin, 2019; Veremyev et al., 2019), and used network structures to reveal characteristics of discourse content. Finally, network methods are also used to tackle the integrated domain of collaborative discourse. Socio-semantic network analysis is one promising approach that integrates semantic features of discourse with social networks (Oshima et al., 2012). The aforementioned KBDeX examines collaborative discourse by combining social and semantic aspects of discourse. It constructs three networks—a student network, a discourse network, and a word network—based on word co-occurrences and uses centrality measures to characterize network positions of students and words. These network approaches have shown promise in capturing various aspects of the nuanced discourse process.

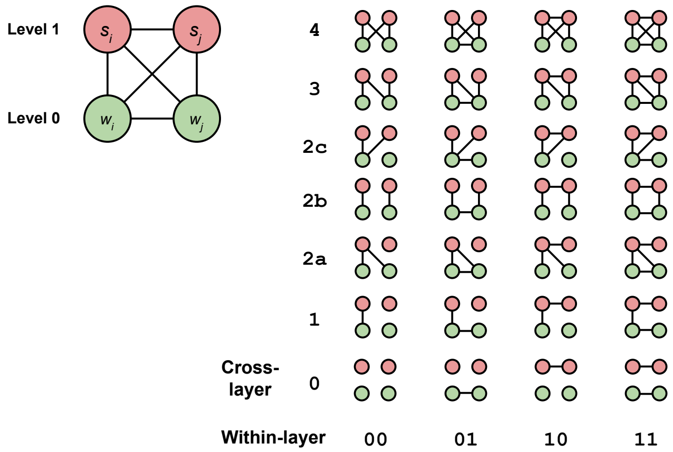
However, there is still a need for integrative approaches to analyzing collaborative discourse—and knowledge-building discourse in particular. To mitigate this challenge, we propose to model knowledge-building discourse as a *socio-semantic network* (SSN) (Basov et al., 2020) and the use theoretically informed *network motifs* in the SSN to characterize knowledge-building discourse. This framework of SSN motif analysis is conceptually grounded in a dual emphasis on social and cognitive/semantic structures in knowledge-building discourse. Guided by this framework, the actors (learners) and semantic entities (words), along with their connections, are modeled as nodes and edges in a *two-mode, dual-layer* socio-semantic network. In contrast with unimode, single-layer networks (e.g., social networks of students, word co-occurrence networks), the dual-layer network approach attempts to treat discourse as a socio-semantic system and maintain the socio-semantic relations in data analysis. With the dual-layer network, we draw on advances in network science to examine the socio-semantic network based on its *network motifs*—defined as recurring, significant patterns of interconnections in the network (Milo et al., 2004). These network motifs capture local network structures that govern behaviors of a network, and have been applied to complex biological, technological, and social networks (Milo et al., 2004). In comparison with traditional SNA measures, such as density and centralization, network motifs capture more subtle network patterns in discourse data and consider the social and semantic properties in tandem. The framework situates network motif analysis in the context of knowledge-building discourse and provides a fresh approach to characterizing discourse using SSN motifs. So far as we know, this paper is the first attempt to apply network motif analysis in CSCL and the Learning Sciences. In the following sections, we introduce the framework SSN motif analysis for knowledge-building discourse and illustrate its application through a case study. We then present key findings and discuss implications and next steps.

The Socio-Semantic Motif Analysis Framework for Knowledge Building

Knowledge building emphasizes students taking collective responsibility over their knowledge as they work together to improve ideas. According to Scardamamlia (2002), "the discourse of knowledge building communities results in more than the sharing of knowledge; the knowledge itself is refined and transformed through the discursive practices of the community—practices that have the advancement of knowledge as their explicit goal" (pp. 11-12). This idea is congruent with socio-cognitive and socio-cultural views of learning that are shared among many CSCL theories (Stahl & Hakkarainen, 2021). One important feature of knowledge-building discourse is its clear emphasis on emergent, progressive knowledge advancement that is not (solely) driven by external curricular goals but ideas produced by learners as they participate in discourse processes (Chen & Hong, 2016). However, supporting such discourse is challenging because it requires a fundamental transformation from classroom discourse that centers on curricular ideas to knowledge-building discourse that values students' epistemic agency and emergent collaboration. The challenge with fostering knowledge-building discourse has motivated scholars to develop sophisticated scaffolding strategies (Chen & Hong, 2016) as well as analytic tools such as the Knowledge Connections Analyzer (KCA) and Knowledge Building Discourse Analyzer (KBDeX) to support students' reflective assessment (Oshima et al., 2012; Yang et al., 2021).

The framework of socio-semantic network (SSN) motif analysis proposed in this paper is grounded in the view of knowledge-building discourse as an emergent socio-semantic system. In knowledge-building discourse, a community of learners are engaged in producing, sharing, and improving ideas. Both social interaction and shared attention to cognitive content are essential for knowledge-building discourse. If a group of learners are simply sharing content with little semantic overlap, idea improvement and "rise-aboves" (i.e., higher-level structures of ideas; Scardamalia (2002)) required by knowledge building are unlikely to happen. By the same token, if learners have minimal social interaction, the discourse may not give rise to community knowledge, diverse ideas, or symmetric knowledge advancement knowledge building aspires to achieve. Thus, the socio-semantic duality in student discourse, explored in various settings including history and scientific communication (Basov et al., 2020), is essential for knowledge-building discourse.

Figure 1
An SSN motif, and a classification of motifs in collaborative discourse. An SSN motif consists of two words on Level 0 (w_i and w_j), two students on Level 1 (s_i and s_j), and six possible links among them. The classification system is defined based on the knowledge-building discourse process. A class is generally anticipated to start from motifs with lower connectivity in the lower-left corner and move towards more connected motifs in the upper-right corner.



Network motifs, a concept from network science, is defined as recurring, significant patterns of interconnections in the network (Milo et al., 2004). Here, network *pattern* means small induced subgraph from the whole graph/network; *recurrence* of a motif indicates its high frequency in the network; *significance* of a motif means it appears significantly more frequently than expected, usually in comparison with random graphs simulated using a set of parameters (e.g., number of nodes, edges, density). Network motifs have been widely used to examine a variety of large-scale networks including biological, technological, infrastructural, and social networks (Milo et al., 2004). In environmental science, two-layer network motifs are particularly meaningful for the investigation of socio-ecological systems that involve social actors (such as tribes, fishing vessels) and ecological resources (e.g., forests, fishing sites) (Bodin & Tengö, 2012).

Our framework applies the network motifs approach to knowledge-building dialogues that are modeled as socio-semantic networks. In a socio-semantic network (SSN), network motifs are minimal sets of social and semantic entities that are basic building blocks of the socio-semantic network. In our framework, each SSN motif consists of two students on Level 1 and two words on Level 0, as well as six potential links among them (see Fig. 1, *left*). As knowledge-building discourse emphasizes the production of knowledge by building on each other's ideas, the SSN motif incorporates three types of links: (a) build-on (or reply) connections among learners on Level 1; (b) logs of writing from learners on Level 1 to words on Level 0; and (c) associations of words in the discourse space. By extracting all possible SSN motifs from a socio-semantic network, this framework attempts to characterize discourse using the frequency and significance of these SSN motifs (more details below). It is hoped that the SSN motifs would provide rich indicators about the modeled discourse as a socio-semantic system.

Situating SSN Motifs in Knowledge Building Discourse

For these SSN motifs to be meaningful—and potentially useful and actionable—they need to be grounded in conceptual understanding of discourse processes. Underlying the SSN motifs is a (simplified) network process of knowledge-building discourse. Following the KB design principles (Scardamalia, 2002), discourse in a knowledge building classroom typically starts from learners sharing their initial ideas about authentic problems they and their community care about. At this stage, learners author posts quite independently and the connections among most words are yet to be established. In other words, while learner–word connections are created in this phase, learner–learner and word–word connections remain sparse. After the initial postings, more social interactions are expected to take place. If learner discourse is transactive—meaning new posts building on prior instances of reasoning (Rosé et al., 2017)—learners' shared attention to words would also increase as they develop overlapping written vocabularies. Concepts are also discussed in relation to each other, leading words that are previously unconnected to appear in the same posts. If the discourse is expansive and generative, learners would introduce new words, creating new opportunities for the peers to "catch up." Idea improvement in knowledge-building discourse depends on the intricate network formation process in a socio-semantic system, with ties formed on different layers—social, semantic, and between them—feeding each other.

The described network process gives meanings to possible SSN motifs. As illustrated in Fig. 1 (*right*), we expect productive knowledge-building discourse to generally move from less sophisticated SSN motifs that are sparsely connected to more sophisticated SSN motifs that are densely connected. In this framework, the SSN motifs are organized by the presence of ties on different layers (see Fig. 1). Less sophisticated discourse would contain more sparsely connected SSN motifs, such as totally separate social and semantic systems (00,0), and unconnected learners producing different words (00,2a), and unconnected learners producing the same word (00,2c). In the initial phase of discourse, SSN motifs with a within-layer code of 00—indicating no links on either the social or semantic layer—may be frequent and significant. As discourse progresses, connections among terms are examined as learners collaborate with one another, giving rise to SSN motifs from the 01 and 10 columns. As knowledge is advanced symmetrically in the community, more SSN motifs from the upper 11 column are expected to emerge in discourse, demonstrating shared interest in interconnected terms. In the most sophisticated cases of knowledge-building discourse, we expect to observe more cases of the fully connected SSN motif, (11,4), where a pair of learners engage with two words that are highly connected in discourse. This expectation does not mean that structurally simple motifs are unimportant; rather, we hypothesize that as discourse progresses less sophisticated motifs will evolve into more sophisticated ones, while new instances of less sophisticated motifs would continue to occur, creating new opportunities for improvement. What is important is the continual entanglement between the social and semantic layers, with connections created on one layer fueling new connections on the other.

Modeling Knowledge-Building Discourse as Socio-Semantic Networks

After situating the SSN motif approach in knowledge-building discourse, we need to further operationalize the network model based on discourse data. It is first important to recognize that any modeling is retaining certain information and discarding the rest. Modeling knowledge-building discourse as socio-semantic networks involves a number of analytical decisions that need careful consideration and an iterative process to refine the model based on theoretical, pedagogical, and technological considerations. In our framework, we propose an iterative heuristic approach to network construction explained below.

The first step is to define a theoretically informed and pedagogical aligned network model. In knowledge-building discourse as socio-semantic networks, there is probably less dispute on including learners and words (or word stems) as two layers of the network. However, there are a number of questions about their linkages. For edges linking learners, we may ask: *Do we treat interaction events as edges? Do we maintain the directionality of these edges? Should we only maintain mutual edges? Do we add a weight to an edge? Should we trim edges based on a weight threshold?* For edges linking learners with words, we could ask: *Are these edges based on receptive or productive vocabulary, which are respectively based on reading and writing behaviors? Do we filter learner–word edges based on a threshold? For the word layer, we may ask: Which words should be represented in the network, a predefined list or a certain number of high-frequency words in the discourse? Do we remove dominant words? How is the word–word connection defined given words can be associated in different ways, such as co-occurrence, free association, semantic relationship, or machine-learned similarity?*

In the case study presented in the next section, the top 100 high frequency words (after removing stopwords) which have appeared for minimally 5 times were incorporated in the socio-semantic network. For these learner–word edges, only writing behaviors were considered; in addition, a threshold applied to the learner–word edges, only maintaining edges with a weight greater than 2. In terms of edges among learners, directed edges between two learners were converted to an undirected edge for simplicity. Edges among words were created based on the φ correlation coefficient computed based on their occurrences in discussion posts; the φ coefficient is similar to the Pearson correlation coefficient in its interpretation and word–word links with $\varphi < .4$ (indicating less than moderate correlation) were removed. These analytical decisions were made based on what is considered conceptually important and contextually appropriate. When applying this framework, it is sensible to experiment with multiple sets of parameters to observe fluctuation across different configurations before choosing the final configuration.

Computing SSN Motifs and their Significance Profiles

After the socio-semantic networks were constructed, we compute the SSN motifs in each network. The first step is to count the occurrences of SSN motifs in the empirical network using the motifr R package (version 0.5.0). The motif frequencies could serve as initial indicators of the discourse's motif profile. Next, to examine the significance of the motif frequencies, we generate a number of refined Erdos-Rényi random graphs which contain randomized edges but the same number of nodes and edges on each layer of the SSN. These random graphs serve as the baseline or null model. Using motifr, we compute the motif profile of each random graph. After establishing the baseline, we then compare the real network's motif profile with those of the random graphs. A Z-score is calculated for each SSN motif using the following equation:

$$\mathrm{Z_i} = \frac{\mathrm{N_i^{real}} - \overline{\mathrm{N}}_\mathrm{i}^{\mathrm{rand}}}{\mathrm{std}\left(\mathrm{N_i^{rand}}\right)}$$

where N_i^{real} stands for the frequency of $\mathrm{motif_i}$ in the real network, N_i^{rand} is the frequency of $motif_i$ in a randomized network. To account for the effect of network size, the Z-score is further normalized to a normalized Z-score, named motif *significance profile* (SP), using the following equation:

$$SP_i = \frac{Z_i}{\sqrt{\sum_j Z_j^2}}$$

The SP score ranges from -1 to 1, and indicates the under- and over-representation of each SSN motif.

For a socio-semantic network, a vector of SP scores of all SSN motifs form this network's motif signature. This motif signature provides a means to characterize knowledge-building discourse by the representation of SSN motifs that are grounded in knowledge-building theory and discourse processes. This framework of SSN motif analysis distinguishes itself from traditional network measures (such as *density* and *average degree*) to capture socio-semantic patterns that are critical for knowledge-building discourse. This framework also provides a computational approach to incorporating discourse content in network analysis without going through laborious content analysis. These SSN motifs provide a formalized way to capture how learners as social actors are connected with words—which are building blocks of discourse content. At this point, the SSN motif signature generated by the framework intends to be descriptive (instead of evaluative) of knowledge-building discourse.

Case Study: Examining Student Discourse Using the SSN Motif Framework

Context and Data

To illustrate the application of the SSN motifs framework, we examine knowledge-building discourse in two 9th grade classes. This case study drew on a secondary dataset generated from design-based research project that aimed to connect discourse in 9th grade science with public discourse in society (Chen et al., 2020). A classroom intervention was conducted in five 9th-grade science classes of an urban public school in the US. Students in these classes were taught by the same teacher who had been using Knowledge Forum for more than five years. During the pilot, the classes worked on two curriculum units about *Energy* and *Elements*. This pilot spanned from February to May and was naturally dived by these two units into two phases. With the Green New Deal (GND) trending in the news, the teacher situated student work within public discourse on this highly contested topic. During the pilot, students were asked to (a) identify authentic knowledge problems relevant to GND, (b) formulate problems in groups, (c) annotate information on the web while seeking to understand and solve these problems, and (d) engage in evidentiary reasoning on Knowledge Forum using annotations. Students used Knowledge Forum to record their ideas and improve them collaboratively. Prior qualitative content analysis has examined the ways in which students navigated their discourse spaces and constructed understanding based on collective sensemaking of public discourse (Chen et al., 2020). This paper draws on two contrasting classes that both showed progress across two phases while also demonstrating different discourse patterns. In particular, Class B engaged with more web resources, showed more intensive collaboration, and achieved greater progress on related topics.

Data analyzed in this study included 320 Knowledge Forum posts created by students from two classes. Table 1 provides descriptive statistics of the dataset. As shown in the table, two classes had equivalent number of students who produced comparable number of Knowledge Forum posts. Class A wrote less but longer posts.

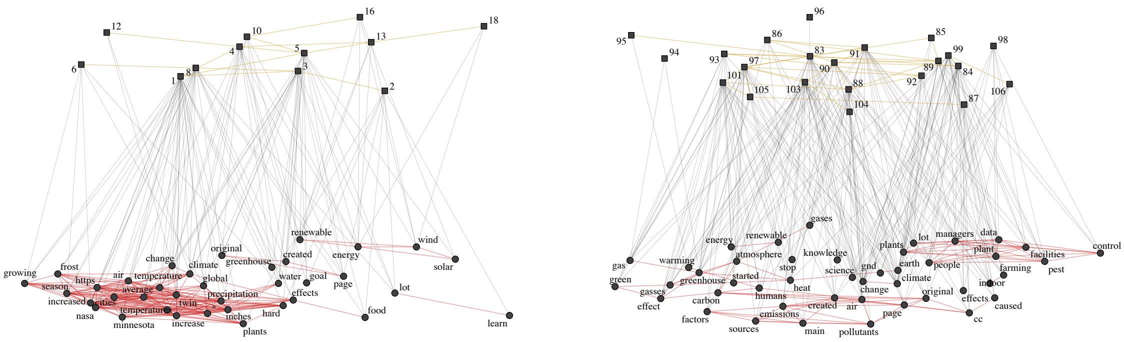
Analysis

Following the SSN motif framework described in Section 2.2, we constructed a socio-semantic network for each discourse phase of each class. Fig. 2 presents example networks created from two classes during Phase 1. In each network, the upper layer displays the undirected interaction network of students and the lower layer is made of the high frequency words generated from students' Knowledge Forum discourse in the particular phase. The links between a student and a word illustrates that the word was mentioned at least twice by the student during the phase. The link between two words was created when their φ correlation coefficient was greater than .04, the threshold of moderate correlation. After the networks were constructed, we applied the computational procedures described in Section 2.3 to these networks by first counting the frequency of SSN motifs and then computing their significance profiles by comparing with 1,000 Erdos-Rényi random graphs.

Table 1
Descriptive Statistics of Discourse Data

Class	Phase	Posts	Interactions	Average words per post
Class A	1	68	33	77.2
(n=22)	2	72	38	102.0
Class B	1	97	40	65.4
(n=26)	2	83	35	60.8

Figure 2
The socio-semantic network of learners and words of two classes, A and B, during Phase 1. Note: Unconnected network components are not shown in the visualizations due to a limitation with the visualization tool.



Findings

Fig. 3 provides an overview of the motif signature of two classes during two discourse phases. The motif signature is organized as a 4 × 6 matrix in the same way as Fig. 1. Each cell in the matrix presents the significance profile (SP) value of the corresponding network motif. For instance, the (11,4) motif had a SP score of 0.64 in Class B during Phase 1 and increased to 0.74 during Phase 2.

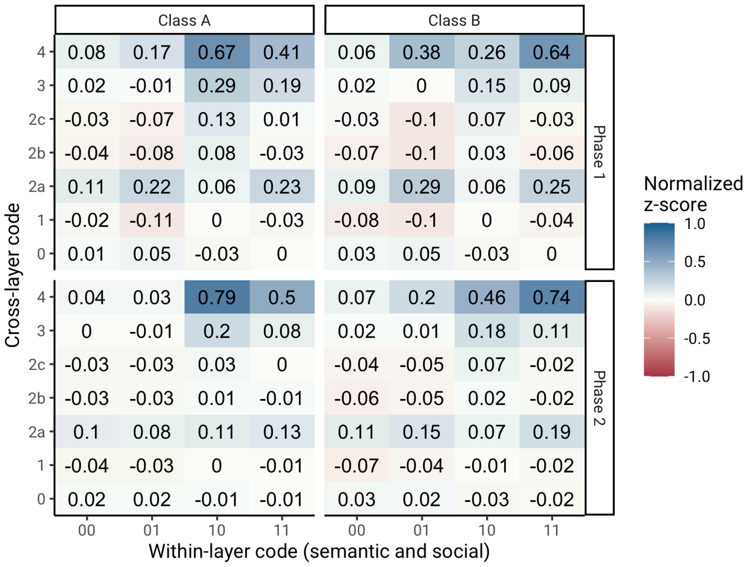
As demonstrated by Fig 3, in general SSN motifs in the upper-right corner were more over-represented in both classes, while motifs with less connectivity in the lower-left corner tended to have negative or small SP values. This finding indicated reasonably sufficient knowledge-building discourse in both classes.

When comparing across two phases, the SP score of (11,4) increased in both classes, from 0.41 to 0.50 in Class A and from 0.64 to 0.74 in Class B. The SP score of (10,4) similarly increased in both classes, indicating the existence of social connection when two learners had shared access to two words. These increases were mostly accompanied by decreases in other adjacent motifs such as (01,4), (10,3), (11,3), demonstrating a transition from less connected motifs to more sophisticated ones across two phases. In addition, the (01,2a) and (11,2a) motifs were above 0.20 in both classes during Phase 1 and decreased to less than 0.20 in Phase 2. Given these two (,2a) motifs indicate exclusive (i.e., non-shared) access to two related words, their decreases indicated more shared vocabulary among learners.

Finally, when comparing two classes, the SP score of (11,4) was higher in Class B while Class A had more significant (10,4) and (10,3). While the SP score of both (11,4) and (10,4) improved in Class A, there was still much room to convert these (10,4) and (10,3) motifs to the fully connected (11,4) motif. This gap could suggest Class A to inspect connections among frequent terms in their discourse. In contrast, Class B demonstrated highly significant (11,4) motifs in both phases, as well as progress on converting the adjacent motifs (such as (01,4)) to more connected motifs in the upper-right corner.

In conclusion, based on the SSN motif analysis, knowledge-building discourse in both classes showed considerable socio-semantic connectivity in both phases. Such connectivity increased across two phases, while Class A still had room for improvement by building connections among high-frequency terms in their discourse.

Figure 3
SSN motif signatures of two classes. Each discourse segment, defined by Class × Phase, is represented by one 4×6 matrix. In each matrix, the network motifs are organized in the same way as in Fig. 1. Each cell in the matrix shows the corresponding motif's significance profile (SP) value (i.e., normalized Z-score). The SP value is also visualized as the cell's background color, on a red–blue color scheme.



Discussion and Implications

This paper proposes a nascent socio-semantic network (SSN) motifs framework for the analysis of knowledge-building discourse. This framework is grounded in the Knowledge Building theory, inspired by advances in network science, and motivated by a need for more integrated approaches to investigating learner dialogues. Using network motif analysis that has been widely applied to complex networks (Milo et al., 2004) but is virtually non-existent in CSCL and the Learning Sciences, the SSN motifs framework introduces a new toolkit not only for the analysis of knowledge-building discourse but also for collaborative discourse in general. The two-layer, socio-semantic network approach enables the investigation of the socio-semantic duality critical for effective discourse (Basov et al., 2020). Overall, we believe the introduced framework of SSN motif analysis makes a strong contribution to CSCL and the Learning Sciences.

In this paper, we applied the framework to knowledge-building discourse in two classes. Results showed similar characteristics of discourse in these classes as well as distinct motif signatures in different discourse segments (divided by class and phase). Following the experimentation, we envision the framework to be used in multiple ways. First, as demonstrated by Fig. 3, the SSN motif signature can provide a rich summary of knowledge-building discourse—or different discourse segments defined by time, activities, or discourse spaces. In comparison with traditional descriptive statistics and SNA metrics, the SSN motifs can provide nuanced information about discourse, which can be used to inform student reflection or pedagogical actions. For example, the over-representation of lower-left motifs in Fig. 1 calls for more socially and semantically cohesive discourse; the over-representation of (,2a) motifs could indicate the need to promote shared attention to key terms among learners; the extremely high significance of the (10,4) motif and moderate significance of (11,4) (shown in Class A of the study) hints a need to foster richer connections among terms. Of course, in which ways such information can be used to inform teaching is to be further designed. Second, one area we could further explore is to identify "critical gaps" using motif analysis to identify high impact links that could create a large number of sophisticated motifs. If (11,4) is important for knowledge building, identifying high impact links that can lead to the creation of (11,4) motifs could suggest actions such as connecting two students or asking the class to discuss possible connections between two terms. Finally, the framework is generic enough to be adapted to different discourse contexts. The motif classification system presented in Fig. 1 was developed for knowledge-building discourse but can be revised for a different context. The analytical decisions described in Section 2.3 could (or should) be revised to fit a new context.

In conclusion, the proposed framework aspires to capture important socio-cognitive dynamics in knowledge-building discourse using socio-semantic networks. This paper makes the critical step to establish a proof-of-concept. Future work will further refine the motif classification system, apply the framework to other discourse contexts, combine SSN motif analysis with other analytical methods (such natural language processing), and evaluate the framework's potential in supporting reflective assessment and pedagogical interventions.

References

Basov, N., Breiger, R., & Hellsten, I. (2020). Socio-semantic and other dualities. *Poetics*, *78*, 101433.

Bodin, Ö, & Tengö, M. (2012). Disentangling intangible social–ecological systems. *Global Environmental Change*, *22*(2), 430–439. https://doi.org/10.1016/j.gloenvcha.2012.01.005

Burtis, P. J. (1998). *Analytic Toolkit for Knowledge Forum*. Centre for Applied Cognitive Science, The Ontario Institute for Studies in Education/University of Toronto.

Chen, B., Chang, Y.-H., & Groos, D. (2020). Bridging Public Discourse and Knowledge Building Discourse in Science Classrooms with the IdeaMagnets Tool. In M. Gresalfi & I. S. Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020, Volume 2* (pp. 863–864). International Society of the Learning Sciences.

Chen, B., & Hong, H.-Y. (2016). Schools as knowledge-building organizations: Thirty years of design research. *Educational psychologist*, *51*(2), 266–288. https://doi.org/10.1080/00461520.2016.1175306

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, *11*(3), 224–238.

Kovanović, V., Joksimović, S., Gašević, D., Hatala, M., & Siemens, G. (2017). Content analytics: The definition, scope, and an overview of published research. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 81–100). SoLAR.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., & Alon, U. (2004). Superfamilies of Evolved and Designed Networks. *Science*, *303*(5663), 1538–1542.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903–921. https://doi.org/10.1007/s11423-012-9265-2

Paranyushkin, D. (2019). InfraNodus: Generating Insight Using Text Network Analysis. *The World Wide Web Conference*, 3584–3589. https://doi.org/10.1145/3308558.3314123

Resendes, M., Scardamalia, M., Bereiter, C., Chen, B., & Halewood, C. (2015). Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 309–336. https://doi.org/10.1007/s11412-015-9219-x

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. *Computer Supported Collaborative Learning* (pp. 69–97). Springer, Berlin, Heidelberg.

Rosé, C. P., Howley, I., Wen, M., Yang, D., & Ferschke, O. (2017). Assessment of Discussion in Learning Contexts. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative Assessment of Collaboration* (pp. 81–94). Springer International Publishing. https://doi.org/10.1007/978-3-319-33261-1 6

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Scardamalia, M. (2004). CSILE/Knowledge Forum. In A. Kovalchick & K. Dawson (Eds.), *Education and technology: An encyclopedia* (pp. 183–192). ABC-CLIO.

Shaffer, D. W. (2017). *Quantitative Ethnography*. Cathcart Press.

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In U. Cress, J. Oshima, C. Rose´, & A. Wise (Eds.), *International handbook of computer-supported collaborative learning*. Springer.

Teplovs, C., Donoahue, Z., Scardamalia, M., & Philip, D. (2007). Tools for concurrent, embedded, and transformative assessment of knowledge building processes and progress. *Proceedings of the 8th Iternational Conference on Computer Supported Collaborative Learning*, 721–723.

Veremyev, A., Semenov, A., Pasiliao, E. L., & Boginski, V. (2019). Graph-based exploration and clustering analysis of semantic spaces. *Applied Network Science*, *4*(1), 1–26.

Yang, Y., van Aalst, J., & Chan, C. (2021). Examining Online Discourse Using the Knowledge Connection Analyzer Framework and Collaborative Tools in Knowledge Building. *Sustainability*, *13*(14), 8045.

Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences*, *27*(3), 390–430.