Socio-Semantic Network Motifs Framework for Discourse Analysis

Bodong Chen
Xinran Zhu
Hong Shui
chenbd@umn.edu
zhu00323@umn.edu
shui0003@umn.edu
University of Minnesota
Minneapolis, MN, USA

ABSTRACT

Effective collaborative discourse requires both cognitive and social engagement of students. To investigate complex socio-cognitive dynamics in collaborative discourse, this paper proposes to model collaborative discourse as a socio-semantic network (SSN) and then use *network motifs* – defined as recurring, significant subgraphs – to characterize the network and hence the discourse. To demonstrate the utility of our SSN motifs framework, we applied it to a sample dataset. While more work needs to be done, the SSN motifs framework shows promise as a novel, theoretically informed approach to discourse analysis.

CCS CONCEPTS

• Information systems → Social networks; **• Human-centered computing → Social tagging systems**; **• Applied computing → Collaborative learning**.

KEYWORDS

networks, collaboration, discourse, two-mode networks

ACM Reference Format:
Bodong Chen, Xinran Zhu, and Hong Shui. 2022. Socio-Semantic Network Motifs Framework for Discourse Analysis. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3506860.3506893

1 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," they also learn collaboratively by constructing a shared problem space, building on one another, and creating shared knowledge artifacts [15, 17]. *Collaborative discourse* is one genre of CSCL designs inspired by the socio-cognitive paradigm of learning. 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.

The learning analytics community has a strong interest in collaboration in general, and collaborative discourse in particular. Since the inception of learning analytics, colleagues have been applying a range of methods to examine collaborative discourse, with a particular focus on the social and cognitive domains of learning [3, 16]. Social network analysis (SNA) – both network metrics and visualizations – are broadly used to investigate social interaction in collaborative discourse [5]. Computational content analysis is also used to examine cognitive content in collaborative discourse [9]. Interesting efforts are also made to bridge qualitative coding of discourse data and quantitative summarization of discourse patterns [18].

However, the community is still in need of more integrative approaches to analyzing collaborative discourse. To mitigate this challenge, we propose to model collaborative discourse as a socio-semantic network (SSN) [1] and then use *network motifs* in the SSN to characterize collaborative discourse. This SSN motifs framework is conceptually grounded in a dual emphasis on social and cognitive/semantic structures in collaborative 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, co-occurrence networks of words), the dual-layer network approach attempts to treat discourse as socio-semantic systems 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 [10]. These network motifs capture local network structures that govern behaviors of a network, and have been applied to complex biological, technological, and social networks [10]. In comparison with traditional SNA measures, such as density and centralization, network motifs capture more subtle network patterns. Overall, the

LAK22, March 21–25, 2022, Online, USA

ACM ISBN 978-1-4503-9573-1/22/03…$15.00
https://doi.org/10.1145/3506860.3506893

SSN motifs framework situates network motif analysis in the context of collaborative discourse and provides a fresh approach to characterizing collaborative discourse using SSN motifs. In the following sections, we first review related work. We then introduce the SSN motifs framework for collaborative discourse and illustrate its application using a secondary dataset. We conclude this paper by discussing key findings, implications, and next steps.

2 RELATED WORK

2.1 Collaborative Discourse

Language is an important tool for learning [22]. Discourse, a direct exchange between humans who can attribute intentionality and understanding to one another, is the foundational act of language and is widely accepted to be an important tool for learning.

Collaborative discourse in CSCL research is an attempt to leverage learners' interpersonal communication and intersubject meaning-making to achieve learning goals beyond each individual [19]. Influenced by socio-cognitive views of learning, collaborative discourse aims to leverage both cognitive and social processes for learners to engage in activities such as articulation, explanation, questioning, and knowledge co-construction [20]. Take the ICAP framework for example. It identifies four increasingly productive kinds of learning processes including *passive*, *active*, *constructive*, and *interactive* processes [4]. According to ICAP, in comparison with *passive* learning (e.g., listening to lectures) and *active* learning (e.g., underlining text sentences), learning in *constructive* conditions requires the use of prior knowledge to interpreting information, whereas learning in *interactive* conditions further involves collaboratively co-constructing solutions or elaborating upon one another's ideas [4]. Sophisticated collaborative discourse in the advanced interactive condition implies complex dynamics of social and cognitive processes.

2.2 Analysis of Collaborative Discourse

Collaborative discourse is a multi-faceted phenomenon that can benefit from diverse analytic approaches. Generally speaking, there are at least three analytical domains of interest to collaborative discourse – cognitive, social, and integrated domains [3]. Questions in different domains lend themselves to different analytical methods.

For the cognitive domain, content analysis is often applied to discourse data to examine specific constructs of cognition. In collaborative knowledge building, for instance, different discourse moves – e.g., asking a question, proposing an explanation, obtaining information – are identified in textual data to evaluate the richness of discourse [14].

The social domain of collaborative discourse focuses on constructs related to group dynamics, coordination, and affective factors [3]. These constructs, such as social solicitation [25] or expressed emotions [26], can be examined using content analysis as well. Also popular in the analysis of the social domain is to apply Social Network Analysis (SNA) [23] to reveal social structures in student interaction based on network visualizations and metrics such as centrality, density, and modularity [13].

The integrated domain considers multiple cognitive and social constructs in tandem. The notion of transactive discussion, or *transactivity* [17], is one example phenomenon in the integrated domain. Transactivity highlights the connection between reasoning (in the cognitive domain) and group dynamics (in the social domain) in that transactive contributions need to build on prior instances of reasoning in the discussion. Recent work in learning analytics has highlighted the prospect of examining cognitive content and social positioning in online discourse [6, 7].

2.2.1 *Network Approaches to Analyzing Collaborative Discourse.*

Network analysis methods are widely used to examine collaborative discourse in cognitive, social, and integrated domains. Questions concerning the social aspect of discourse lend themselves to SNA. Social network centrality measures (e.g., degree, closeness, and betweenness) are used to evaluate a student's social position in a class and then correlate with students' sense of community [5]. SNA is also popularly used to analyze student interactions (*cf.* social relations) in discussion forums, using network-level measures such as network density and average degree to characterize different interaction networks [e.g., 24].

Network methods are also used to analyze and represent cognitive content. For example, content entities such as words and artifacts can be studied as networks. Prior work has leveraged different types of word relations, such as co-occurrence, word sequence, word embedding-based similarity [12, 21], and used network structures to reveal characteristics of discourse content. The nature of discourse content, such as being dispersed, biased, focused, or diversifed, can be characterized based on the structure of word networks [12].

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 [11]. The Knowledge Building Discourse Explorer (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. Similarly, actor-artifact networks are also used to characterize artifacts and author profiles in Wikipedia editing [8].

Overall, network analysis offers rich opportunities for the analysis of collaborative discourse. However, there are a few serious challenges. First, researchers need to more explicitly address discourse processes and assumptions when constructing networks from discourse data [13]. Second, given the close-knit relationship between cognitive and social aspects of discourse, we need more ways to examine the integrated domain of discourse. Finally, we also need to develop more actionable discourse indicators to improve the impact of discourse analytics. In the next section, we introduce a socio-semantic motifs framework proposed to mitigate these challenges.

3 THE SOCIO-SEMANTIC MOTIFS FRAMEWORK

This framework is based on the basic assumption that collaborative discourse requires multiple learners discussing shared content. Social interaction and shared attention are both essential for collaborative discourse. If learners have no or minimal interactions, the discourse cannot be collaborative. Similarly, if a group of learners

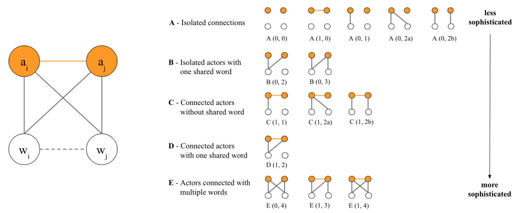

Figure 1: A SSN motif, and a classification of motifs in collaborative discourse. Note: We created the naming system based on the number of edges on each layer. For instance, `A(1,0)` has one edge on the top layer, while `C(1,1)` adds another edge between two layers. `A(0,2a)` and `A(0,2b)` have the same number of between-layer edges but differ in the edge combination.

are simply sharing content without semantic overlap, intersubjective meaning-making necessitated for collaboration is unlikely to happen.

The socio-semantic network (SSN) motifs are minimal sets of social and semantic entities that are basic building blocks of a socio-semantic network. In our framework, each SSN motif consists of two learners and two words, which has six potential links (see Fig. 1). In network science, network motifs have been widely used to examine a variety of networks including biological, technological, infrastructural, and social networks [10]. 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) [2]. Our framework similarly applies two-layer network motifs to collaborative discourse. 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. Our hypothesis is that the frequency of the SSN motifs, as well as their significance relative to null models, would provide rich indicators about the modeled discourse as a socio-semantic system.

3.1 Situating SSN Motifs in Collaborative 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 informed by the collaborative discourse literature (see Sec. 2.1). In the beginning of collaborative discourse, learners would write posts about ideas they are interested in. In this process, learners produce words quite independently, leading to little shared attention to words. After the initial postings, more social interactions are expected to take place, leading to increased social connections among learners. If learner discourse is indeed transactive, meaninging new posts building on prior instances of reasoning [17], learners' shared attention to words would increase as they develop overlapping written vocabularies. If the discourse is expansive and generative, learners would also expand to cover new words, creating new opportunities for the peers to "catch up."

The described network process gives meanings to possible SSN motifs. As illustrated in Fig. 1, we expect healthy collaborative discourse to generally move from less sophisticated SSN motifs that are sparsely connected to more sophisticated SSN motifs that are densely connected. This expectation does not mean structurally simple motifs are unimportant; rather, we hypothesize that as discourse progresses more sophisticated motifs would emerge besides less sophisticated ones. In this framework, we do not consider possible word–word associations, and propose five levels of SSN motifs (see Fig. 1). Less sophisticated discourse would contain more sparsely connected SSN motifs, such as totally separate social and semantic systems `A(1,0)`, isolated learners `A(0,0)` and `A(0,1)`, and unconnected learners producing different words `A(0,2b)`. In more sophisticated collaborative discourse, social connections emerge, generating SSN motifs of `B(0,3)`, `C(1,2a)` and `C(1,2b)`. Even more sophisticated cases would include SSN motifs that show increased social contacts also leading to overlapping interest in shared vocabulary, `D(1,2)` and `E(1,3)`. In the most sophisticated cases of collaborative discourse, we expect to observe more cases of fully connected SSN motifs, `E(1,4)`, where pairs of learners engage with shared words.

3.2 Modeling Collaborative Discourse as Socio-semantic Networks

It is first important to recognize that any modeling is retaining certain information and discarding the rest. Modeling collaborative discourse as socio-semantic networks involves a number of analytical decisions that are not always straightforward and require 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. Here we would like to admit upfront that this approach involves necessary simplification that does not distinguish different constructs such as words, semantic entities, and semantic structures.

The first step is to define a theoretically informed and pedagogical aligned network model. In collaborative 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?* When needed, edges among words may also be added in different ways, such as co-occurrence, free association, or similarity based on approaches such as latent semantic analysis or word embedding.

In the case study presented in the next section, the top 100 high frequency words that have appeared for minimally 5 times are incorporated in the socio-semantic network. Edges among words were not considered, meaning the network only contained learner–learner and learner–word edges. 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 or equal to 2. In terms of edges among learners, directed edges between two learners were converted to an undirected edge for simplicity. These analytical decisions were made based on what is considered conceptually important and contextually appropriate. When applying this framework, it is also sensible to experiment with multiple sets of parameters to observe fluctuation across different configurations before choosing the final configuration.

3.3 Computing SSN Motifs and the Significance Profile

After the socio-semantic network is constructed, we compute two-layer SSN motifs. The first step is to count the occurrences of interested 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:

$$Z_i = \frac{N_i^{real} - \overline{N}_i^{rand}}{std(N_i^{rand})} \tag{1}$$

where N_i^{real} stands for the frequency of $motif_i$ 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 on the Z-score, it is further normalized to a normalized Z-score using the following equation:

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

The normalized Z-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 normalized Z-scores of all SSN motifs form this network's motif significance profile. Using the profile, this framework provides a means to characterize collaborative discourse by the representation of SSN motifs that are theoretically informed and pedagogically aligned. This approach distinguishes itself from traditional network measures (such as density and average degree) to capture socio-semantic patterns that are critical for collaborative discourse. This framework also provides a computational approach to incorporating discourse content in network analysis without going through laborious content analysis. The patterns captured by SSN motifs are indeed different from constructs of discourse content (e.g., argumentation). Nonetheless, 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. They intend to be descriptive instead of evaluative of discourse quality or student learning.

4 STUDYING SOCIAL ANNOTATION USING THE SSN MOTIFS FRAMEWORK

To illustrate the application of the SSN motifs framework, we examine collaborative discourse in a small online class. In this section, we first describe the study context and discourse data. We then report findings from the analysis of SSN motifs, as well as the association between motifs and content analysis results.

4.1 Context and Data

This study drew on a secondary dataset generated from an undergraduate online course offered by a large public university in 2020. In this course, students (n = 13) were engaged in reading course materials, participating in weekly virtual meetings, and writing reflective essays. For the reading tasks, the instructor used a web annotation tool named Hypothes.is to support collaborative sensemaking via social annotation. Throughout the semester, students were required to read 1-2 readings each week, annotate each reading, and reply to each other's annotations.

Over one semester, the class generated 478 Hypothes.is annotations and 469 replies in 18 readings across 11 weeks. On average, each reading had 26.6 annotations (*SD* = 2.6) and 26.1 replies (*SD* = 4.0). To assess the quality of student annotations in a prior study, we applied to each annotation a four-level *knowledge construction* coding-scheme comprising (1) Initiation, (2) Exploration, (3) Negotiation, (4) Co-construction. The higher levels indicate students' higher-order thinking skills in areas such as negotiation, synthesis, meta-cognition. The average *knowledge construction* score of all student posts was 2.36 (*SD* = 0.21). Table 1 provides descriptive statistics of student annotations, replies and average knowledge construction score in each reading.

4.2 Findings

4.2.1 Motif profiles of collaborative discourse of readings. Following the procedure described in Section 3.2, we constructed two-layer socio-semantic networks where the upper layer displays the undirected interaction network of students and the lower layer is made of the high frequency words generated from students' written discourse in a given week. The links between a student and a word illustrates that the word was mentioned at least twice in the student's posts about this particular reading. The links between words were not considered. Fig. 2 presents an example socio-semantic network created from discourse around a particular reading.

As described in Sec. 3.3, we then conducted motif analysis by first counting the frequency of SSN motifs and then computing the significance profile of each reading. Table 2 presents the mean and standard deviation of the SSN motifs' significance profile across all readings in this course. Note that a larger positive value indicates significant over-representation in comparison with random graphs, whereas a smaller negative value means significant under-representation. Overall, `E(1,4)` and `A(0,2a)` were the over-represented motifs in this dataset, whereas `A(0,2b)`, `B(0,2)`, and `C(1,2b)` were the most under-represented. Note that `A(0,2a)` means

Table 1: Descriptive Statistics of Annotations, Replies and Knowledge Construction.

Readings	1	2a	2b	3a	3b	4	5	6a	6b	7a	7b	8	9a	9b	10a	10b	11a	11b
Annotations	28	27	32	27	26	29	28	28	28	24	26	29	29	23	22	24	26	22
Replies	25	28	24	27	24	38	27	30	23	28	27	27	26	25	22	21	28	19
Knowledge Construction	1.98 (0.97)	2.27 (0.73)	2.07 (0.85)	2.75 (0.61)	2.30 (0.76)	2.55 (0.64)	2.38 (0.62)	2.59 (0.65)	2.02 (0.76)	2.46 (0.61)	2.19 (0.82)	2.45 (0.74)	2.22 (0.71)	2.40 (0.61)	2.48 (0.85)	2.24 (0.74)	2.56 (0.79)	2.58 (0.77)

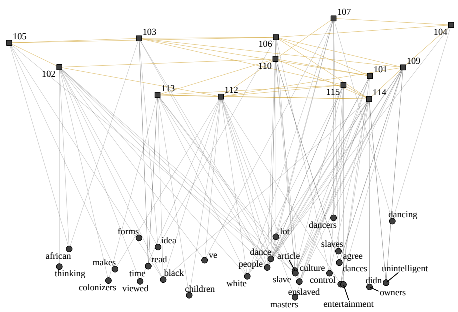

Figure 2: The socio-semantic network of learners and words of Reading #6a

one student is connected to two words but there is no connection between students, whereas E(1,4) indicates students and words are all connected. Our interpretation of them being over-represented is that the frequent occurrence of sophisticated motifs does not always lead to the decrease of less sophisticated motifs. While A(0,2a) is characterized as having isolated connections, it also indicates an interesting situation where one student is pulling the discourse towards new directions and providing opportunities to expand current discourse. Thus, it is insufficient to evaluate discourse based on one SSN motif; instead, it is critical to concurrently consider multiple motifs.

To compare across all readings, Fig. 3 visualizes the SSN motifs profiles of the readings using radar charts. SSN motifs are sorted in each radar chart from less sophisticated motifs (e.g., A(0,0)) to the most sophisticated one (i.e., E(1,4)) in the counter-clockwise direction. This figure first provides an overview of the class' participation network. In general, the SSNs are well connected in this course as indicated by the general over-representation of sophisticated motifs (e.g., around the top right corner) and under-representation of less sophisticated motifs (e.g., around the top and bottom left corners).

It is also possible to zoom into a particular reading. In Fig. 3, each profile shows a distinct polygon yet there are similarities among several readings. For example, Readings 3a, 7a, and 7b show very similar SSN motif structures. They all have most of the sophisticated motifs (e.g., E(1,4), E(1,3)) over-represented and most of the less connected motifs (e.g., A(0,2b)) under-presented. In comparison, Reading 3b has stronger representation of A(1,0), B(0,2), and B(0,3), indicating a different discourse profile. The differences in SSN motifs imply that students had less cohesive discourse around Reading 3b, in comparison with the other three readings.

4.2.2 Correlation between motifs and knowledge co-construction. To examine the extent to which these network motifs were associated with *knowledge construction*, we correlated the normalized Z-score of each SSN network motif with the average knowledge construction score. As presented in Fig. 4, in this particular context, A(0,1) and E(1,4) were positively correlated with the knowledge construction score (Spearman's $\rho > .4$), while A(0,2b) and D(1,2) had weak positive correlation with knowledge construction (Spearman's $\rho > .2$). In contrast, B(0,3), C(1,1), and C(1,2a), and E(0,4) were negatively correlated with knowledge construction (Spearman's $\rho < -.2$). A closer inspection of Table 2 found the average normalized Z-scores of motifs A(0,1), B(0,3), C(1,2a), and E(0,4) close to zero, meaning they were nonsignificant in comparison with the random graph baseline. Therefore, higher Z-scores of E(1,4) and A(0,2b) were associated with higher knowledge construction, whereas higher C(1,1) were linked to lower knowledge construction. This finding indicated the particular importance of E(1,4) and A(0,2b) for facilitating knowledge construction.

5 DISCUSSION AND IMPLICATIONS

This short paper proposes a nascent socio-semantic network (SSN) motifs framework for the analysis of collaborative discourse. This framework is grounded in socio-cognitive learning theories, inspired by advances in network science, and motivated by a need for more integrated approaches to investigating learning dialogues. Using network motif analysis that has been widely applied to complex networks [10] but is virtually non-existent in learning analytics, the SSN motifs framework introduces a new toolkit for discourse analysis. The two-layer, socio-semantic network approach adds additional richness to this framework, enabling the investigation of the socio-semantic duality critical for effective discourse [1]. Overall, we believe the introduced SSN motifs framework makes a solid contribution to learning analytics, discourse analysis, and collaborative learning.

In this paper, we applied the framework to an example discourse dataset. Results showed general characteristics of discourse in the class as well as distinct motif profiles of different discourse segments (divided by reading). Further correlational analysis found some SSN motifs associated with human coding of knowledge construction. Following the experimentation, we envision the framework to be used in several ways. First, as demonstrated by Table 2 and Fig. 3, the framework can provide an overview of 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 evaluate instruction and inform pedagogical actions. For example, the over-representation of Level A motifs calls

Table 2: Mean and Standard Deviation of the significance profiles of all SSN Motifs.

Motif	A(0,0)	A(1,0)	A(0,1)	A(0,2a)	A(0,2b)	B(0,2)	B(0,3)	C(1,1)	C(1,2a)	C(1,2b)	D(1,2)	E(0,4)	E(1,3)	E(1,4)
Mean	0.22	-0.04	0.01	0.33	-0.49	-0.21	-0.07	-0.13	-0.02	-0.21	0.12	0.08	0.18	0.39
SD	0.08	0.15	0.06	0.25	0.20	0.26	0.07	0.03	0.17	0.15	0.15	0.10	0.07	0.09

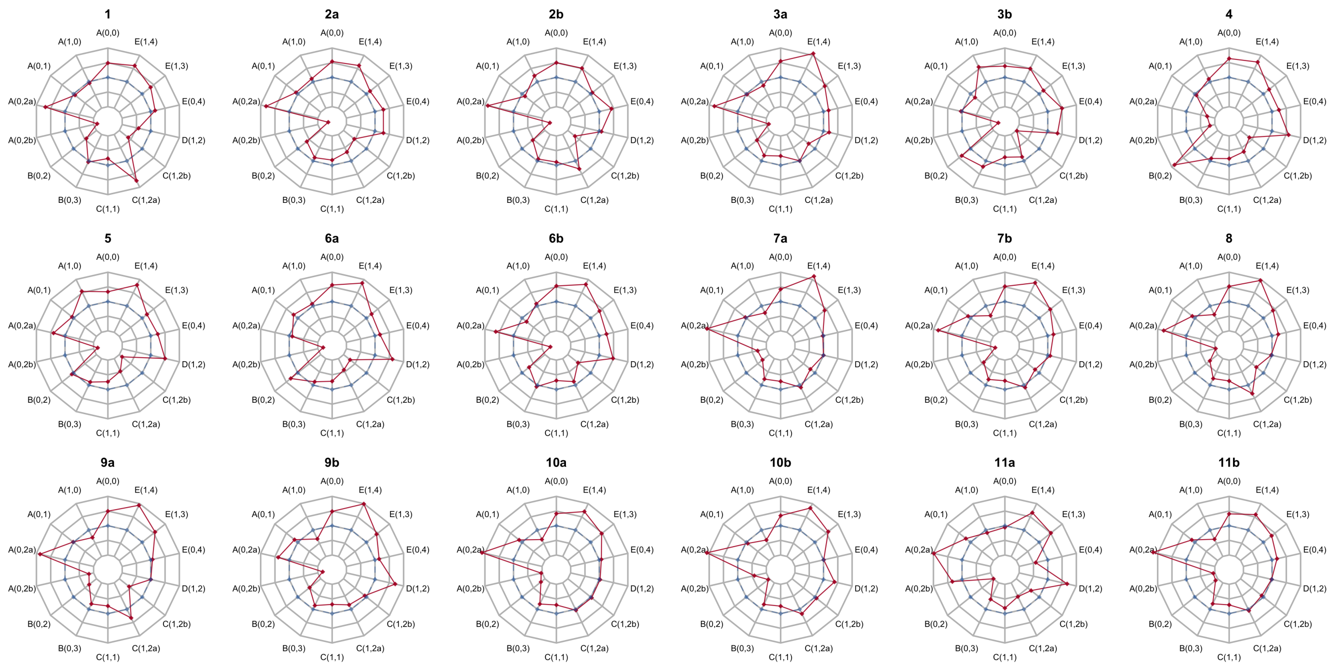

Figure 3: SSN motifs profiles. Each reading is represented by one radar chart. In each chart, the network motifs are sorted from Level A to Level E in the counter clock-wise direction. Each motif's normalized Z-score is depicted by a red dot on the axis. Each axis ranges from -0.5 to 0.5, with the blue line indicating the 0 point. As such, a dot outside of the blue line indicates over-representation of the motif, and vice versa.

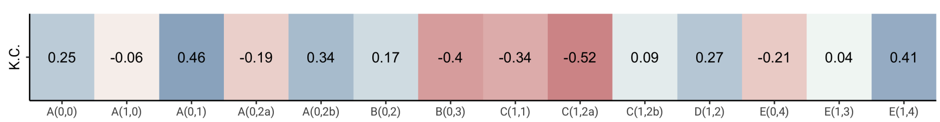

Figure 4: Correlation (Spearman's ρ) between SSN motifs and knowledge construction.

for more socially and semantically cohesive discourse; the under-representation of Levels D and E motifs could indicate the need to facilitate students' shared attention to key concepts. 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 E(1,4) is important for knowledge construction, identifying high impact links related to E(1,4) could suggest actions such as connecting two students. Finally, the framework is generic enough to be adapted to different discourse contexts. The sophistication levels of motifs presented in Fig. 1 can be revised for a different context. Word association can also be added if appropriate, giving rise to even a richer set of SSN motifs to be considered.

In conclusion, the proposed framework aspires to capture important socio-cognitive dynamics in collaborative discourse using socio-semantic networks. This short paper make the initial 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, and integrate pedagogical designs (such as requiring each student to make a certain number of posts and replies) in the baseline model.

REFERENCES

[1] Nikita Basov, Ronald Breiger, and Iina Hellsten. 2020. Socio-Semantic and Other Dualities. *Poetics* 78 (Feb. 2020), 101433. https://doi.org/10.1016/j.poetic.2020.101433

[2] Örjan Bodin and Maria Tengö. 2012. Disentangling Intangible Social–Ecological Systems. *Global Environmental Change* 22, 2 (May 2012), 430–439. https://doi.org/10.1016/j.gloenvcha.2012.01.005

[3] Yuxin Chen, Christopher D. Andrews, Cindy E. Hmelo-Silver, and Cynthia D'Angelo. 2019. Coding Schemes as Lenses on Collaborative Learning. *Information and Learning Sciences* 121, 1/2 (Jan. 2019), 1–18. https://doi.org/10.1108/ILS-08-2019-0079

[4] Michelene T H Chi and Ruth Wylie. 2014. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational psychologist* 49, 4 (Oct. 2014), 219–243. https://doi.org/10.1080/00461520.2014.965823

[5] Shane Dawson. 2008. A Study of the Relationship between Student Social Networks and Sense of Community. *Educational Technology & Society* 11, 3 (2008), 224–238.

[6] Nia M. M. Dowell and Oleksandra Poquet. 2021. SCIP: Combining Group Communication and Interpersonal Positioning to Identify Emergent Roles in Scaled Digital Environments. *Computers in Human Behavior* 119 (June 2021), 106709. https://doi.org/10.1016/j.chb.2021.106709

[7] Dragan Gašević, Srećko Joksimović, Brendan R. Eagan, and David Williamson Shaffer. 2019. SENS: Network Analytics to Combine Social and Cognitive Perspectives of Collaborative Learning. *Computers in Human Behavior* 92 (March 2019), 562–577. https://doi.org/10.1016/j.chb.2018.07.003

[8] Iassen Halatchliyski, Johannes Moskaliuk, Joachim Kimmerle, and Ulrike Cress. 2014. Explaining Authors' Contribution to Pivotal Artifacts during Mass Collaboration in the Wikipedia's Knowledge Base. *International Journal of Computer-Supported Collaborative Learning* 9, 1 (2014), 97–115.

[9] V Kovanović, S Joksimović, D Gašević, M Hatala, and G Siemens. 2017. Content Analytics: The Definition, Scope, and an Overview of Published Research. In *Handbook of Learning Analytics*, Charles Lang, George Siemens, Alyssa Wise, and Dragan Ga\v sević (Eds.). SoLAR, 81–100.

[10] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. 2004. Superfamilies of Evolved and Designed Networks. *Science* 303, 5663 (March 2004), 1538–1542. https://doi.org/10.1126/science.1089167

[11] Jun Oshima, Ritsuko Oshima, and Shunsuke Saruwatari. 2020. Analysis of Students' Ideas and Conceptual Artifacts in Knowledge-Building Discourse. *British Journal of Educational Technology* 51, 4 (July 2020), 1308–1321. https://doi.org/10.1111/bjet.12961

[12] Dmitry Paranyushkin. 2019. InfraNodus: Generating Insight Using Text Network Analysis. In *The World Wide Web Conference (WWW '19)*. ACM, New York, NY, USA, 3584–3589. https://doi.org/10.1145/3308558.3314123

[13] Oleksandra Poquet, Mohammed Saqr, and Bodong Chen. 2021. Recommendations for Network Research in Learning Analytics: To Open a Conversation. In *Proceedings of the NetSciLA2021 Workshop "Using Network Science in Learning Analytics: Building Bridges towards a Common Agenda" (NetSciLA2021) (CEUR Workshop Proceedings, 2868)*, Oleksandra Poquet, Bodong Chen, Mohammed Saqr, and Tobias Hecking (Eds.). Aachen, 34–41.

[14] Monica Resendes, Marlene Scardamalia, Carl Bereiter, Bodong Chen, and Cindy Halewood. 2015. Group-Level Formative Feedback and Metadiscourse. *International Journal of Computer-Supported Collaborative Learning* 10, 3 (Sept. 2015), 309–336. https://doi.org/10.1007/s11412-015-9219-x

[15] Jeremy Roschelle and Stephanie D Teasley. 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning*. Springer, Berlin, Heidelberg, 69–97. https://doi.org/10.1007/978-3-642-85098-1_5

[16] Carolyn P Rose, Dragan Gaesevic, Pierre Dillenbourg, Yohan Jo, Gaurav Tomar, Oliver Ferschke, Gijsbert Erkens, Anouschka van Leeuwen, Jeroen Janssen, Mieke Brekelmans, and Others. 2016. Analytics of Social Processes in Learning Contexts: A Multi-Level Perspective. In *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 1*, C-K Looi, J Polman, U Cress, and P Reimann (Eds.). International Society of the Learning Sciences, Singapore, 24–31.

[17] Carolyn Penstein Rosé, Iris Howley, Miaomiao Wen, Diyi Yang, and Oliver Ferschke. 2017. Assessment of Discussion in Learning Contexts. In *Innovative Assessment of Collaboration*, Alina A. von Davier, Mengxiao Zhu, and Patrick C. Kyllonen (Eds.). Springer International Publishing, Cham, 81–94. https://doi.org/10.1007/978-3-319-33261-1_6

[18] David Williamson Shaffer. 2017. *Quantitative Ethnography.* Cathcart Press.

[19] Gerry Stahl and Kai Hakkarainen. 2021. Theories of CSCL. In *International Handbook of Computer-Supported Collaborative Learning*, Ulrike Cress, Jun Oshima, Carolyn Rosé, and Alyssa Wise (Eds.). Springer, Berlin.

[20] Gerry Stahl, Timothy D Koschmann, and Daniel D Suthers. 2006. Computer-Supported Collaborative Learning: An Historical Perspective. In *Cambridge Handbook of the Learning Sciences*, R Keith Sawyer (Ed.). Cambridge University Press, Cambridge, 409–426.

[21] Alexander Veremyev, Alexander Semenov, Eduardo L. Pasiliao, and Vladimir Boginski. 2019. Graph-Based Exploration and Clustering Analysis of Semantic Spaces. *Applied Network Science* 4, 1 (Dec. 2019), 1–26. https://doi.org/10.1007/s41109-019-0228-y

[22] Lev Semenovich Vygotsky. 1978. *Mind in Society: The Development of Higher Psychological Processes.* Harvard University Press, Cambridge, MA.

[23] Stanley Wasserman and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge, UK.

[24] Alyssa Friend Wise and Yi Cui. 2018. Learning Communities in the Crowd: Characteristics of Content Related Interactions and Social Relationships in MOOC Discussion Forums. *Computers & education* 122 (July 2018), 221–242. https://doi.org/10.1016/j.compedu.2018.03.021

[25] Alyssa Friend Wise, Simone Nicole Hausknecht, and Yuting Zhao. 2014. Attending to Others' Posts in Asynchronous Discussions: Learners' Online "Listening" and Its Relationship to Speaking. *International Journal of Computer-Supported Collaborative Learning* 9, 2 (June 2014), 185–209. https://doi.org/10.1007/s11412-014-9192-9

[26] Gaoxia Zhu, Wanli Xing, Stacy Costa, Marlene Scardamalia, and Bo Pei. 2019. Exploring Emotional and Cognitive Dynamics of Knowledge Building in Grades 1 and 2. *User Modeling and User-Adapted Interaction* 29, 4 (Sept. 2019), 789–820. https://doi.org/10.1007/s11257-019-09241-8