

# CLAto: A Collaborative Learning Analytics Tool for Online Discussion

Feng LIN\*, Alicia Huiling LOH & Chenchen LI  
Singapore University of Social Sciences, Singapore  
\*linfeng@suss.edu.sg

**Abstract:** This paper presents CLAto, a collaborative learning analytics tool that augments existing learning management systems to support students' online discussions. It addresses a gap in current collaborative learning analytics tools, which predominantly focus on behavioral and social dimensions, offer limited support for epistemic processes, and lack holistic integration across these aspects. Grounded in CSCL literature and knowledge-building perspectives, CLAto integrates four analytic constructs: Discussion Participation, Social Interaction, Assignment Progress, and Idea Trajectory. Participation metrics and Social Network analysis capture engagement and relational patterns, while Assignment Progress supports monitoring of task completion. The Idea Trajectory represents the temporal evolution of ideas as interconnected nodes and also includes a Promising Idea feature to surface contributions with high epistemic potential. By integrating behavioral, relational, and epistemic representations, CLAto enables more holistic support for epistemic monitoring and collaborative knowledge construction in online discussions.

**Keywords:** Collaborative learning analytics, CSCL, tool development

## 1. Introduction

Computer-Supported Collaborative Learning (CSCL) examines how people learn together with the support of technologies, emphasizing learning as a collaborative knowledge-construction process grounded in interaction and discourse (Stahl et al., 2014). Asynchronous online discussions are a key form of CSCL, enabling flexible participation across time and place and supporting idea improvement, problem-solving, and deeper understanding (Cheng et al., 2022; Lin et al., 2014; Zhang et al., 2009). However, realizing these benefits is challenging, as high-quality online discussions require sustained instructional orchestration and continuous monitoring of students' collaborative processes, particularly when discussions are large and text-heavy (Kaliisa & Dolonen, 2022). Moreover, effective CSCL depends not only on teacher orchestration but also on students' co-regulation of their collaborative processes (Järvelä et al., 2016). To address these challenges, collaborative learning analytics have emerged to make complex collaborative processes more visible and actionable (Shum & Ferguson, 2012; Wise et al., 2021).

A range of collaborative learning analytics tools has been developed to support online discussion. Many focus on participation metrics and social interaction patterns (e.g., (Chen et al., 2018; Scardamalia, 2004), commonly analyzing quantitative indices (e.g., number of posts and reads) and network relationships among participants. Others employ graphical representations—such as threaded tree structures (Wise et al., 2014; Wise et al., 2013), argument maps (Ouyang et al., 2024), and knowledge-building networks (Oshima et al., 2012) — to support online discussion. Overall, existing tools tend to emphasize behavioral patterns and social interaction, with limited support for the epistemic dimension. Although some tools visualize idea structures, node labels are often derived from keywords, offering limited interpretive depth. Few tools leverage large language models (LLMs) to generate more meaningful representations of ideas. In addition, most tools operate separately from learning management systems (LMS) platforms, limiting integration into instructors' workflows, creating technical and administrative barriers that may hinder sustained classroom adoption.

To address this gap, the present study aims to design and develop a Collaborative Learning Analytics Tool (CLAto) to augment the existing LMS and to holistically visualize students' participation, social interaction, and evolving idea structures. These idea structures are generated and semantically enriched using LLM. CLAto also incorporates practice-oriented features derived from instructors' authentic needs, particularly to support the monitoring and assessment of online discussion assignments. The purpose of this paper is to present the design and development process of CLAto.

## 2. Design and development of CLAto

### 2.1 Design context and data pipeline

In this study, the LMS in use is Canvas. Raw data is retrieved daily via a scheduled API call and stored in an Amazon Web Services (AWS) S3 bucket. A daily cron job then parses, transforms, and joins the data before storing it in a MySQL database. Indexing and caching are used to support efficient querying in CLAto. After processing, another daily cron job identifies new and updated discussion entries and generates concise entry titles using OpenAI's GPT-4o-mini model. CLAto then retrieves and delivers the latest data to end users.

### 2.2 Analytic constructs and interface design

At its current stage, CLAto encompasses four analytic constructs: Discussion Participation, Social Interaction, Assignment Progress, and Idea Trajectory. The tool offers both teacher-facing and student-facing views within one application.

**Discussion Participation.** This section presents multiple representations of students' engagement patterns, including (1) a sortable and downloadable summary table of individual-level interaction metrics (e.g., posts authored, replies received, users interacted with, topics participated in, posts viewed, and likes received); (2) a topic-level bar chart showing the distribution of contributions across discussion threads; and (3) a temporal bar chart depicting weekly posting activity. Together, these visualizations can help instructors and students track participation trends across threads over time.

**Social Interaction.** The social interaction section features both Social Network Analysis (SNA) and a Word Cloud (Figure 1). SNA is a well-established analytical approach in CSCL research (Kaliisa et al., 2022) and is commonly applied in social learning analytics to examine interpersonal relationships and network connectivity (Shum & Ferguson, 2012). As shown in Figure 1a, the SNA visualization represents instructors and students as nodes, with directed edges indicating interactions. Selecting a node highlights it in green, while red edges represent incoming replies and blue edges represent outgoing replies. Interaction density is computed as the proportion of observed to possible directed edges in the network (Lizardo & Jilbert, 2020), with higher values indicating more active interactions. Figure 1b complements this by presenting a word cloud that summarizes frequently used terms, enabling quick identification of key discussion themes.

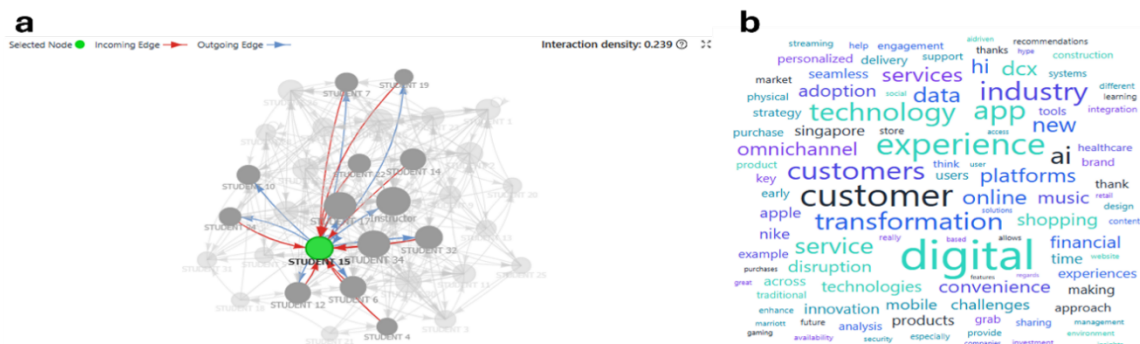


Figure 1. Social Interaction. (a) SNA. (b) Word cloud.

**Assignment Progress.** This feature allows instructors to monitor both the number of students who have completed the discussion assignment and those who have not. They can select topics via a multi-select dropdown and set required post counts per topic to align participation criteria with course objectives. This flexible configuration supports varied expectations across topics and enables effective tracking of assignment completion.

**Idea Trajectory.** The Idea Trajectory feature in CLAtO (Figure 2) visualizes both the structure and temporal development of ideas, showing how individual contributions are connected and developed over the course of a discussion. Unlike the SNA graph, where nodes represent participants, each node in the Idea Trajectory corresponds to a single post.

Just as individual cognition and metacognition operate within a learner, ideas articulated in World 3 (Popper, 1972) within a community also acquire a form of autonomy: they “live” independently and possess their own memory and mechanisms for generating new knowledge. The extent to which this occurs largely depends on how the community cognitively and metacognitively processes the information contributed by its members. Idea Trajectory includes two mechanisms to support the community’s collective metacognition: explication of developing knowledge structure and the identification of promising ideas.

1) *Explication of developing knowledge structure.* Structured knowledge representations have been used to scaffold shared understanding and facilitate the construction of collective knowledge in CSCL (Scardamalia & Bereiter, 2006; Wise et al., 2014; Zheng et al., 2022). By externalizing and organizing ideas, they help establish a joint problem space, support the development of a shared knowledge base, and provide a reference point for participants to coordinate their contributions, monitor collective progress, and regulate the collaborative process (Chen & Lin, 2020). Building on this, CLAtO’s Idea Trajectory visualization makes the evolving knowledge structure visible. It supports students in tracking collective understanding, identifying gaps, and planning contributions, while helping instructors assess discourse depth and provide targeted scaffolding.

Since Canvas discussions do not require post titles, CLAtO uses a LLM (OpenAI) to generate concise semantic labels for each entry, improving interpretability beyond keyword-based approaches (Oshima et al., 2012). In addition, a topic-selection dropdown allows users to switch between a macro-level view of all discussion threads (Figure 2a) and a filtered view of a specific topic (Figure 2b). Clicking a node reveals the original post in a side panel for detailed inspection.



Figure 2. Idea trajectory. (a) Discussion threads overview. (b) Selected discussion thread.

2) *Promising ideas.* To support collective knowledge construction, communities also need to identify and highlight promising ideas that can advance understanding or open new lines of inquiry (Chen, 2017; Wise et al., 2021). The Idea Trajectory includes a Promising Idea function that allows students and instructors to mark posts with a lightbulb icon to signal their value for further exploration. Each post can receive multiple endorsements. This makes epistemically valuable contributions explicit, helping to guide attention and support community-level metacognitive awareness.

### 3. Discussion and conclusion

Grounded in CSCL literature, we developed CLAto, a collaborative learning analytics tool that augments existing LMSs to support online discussions. Unlike many existing tools that focus on limited dimensions, CLAto integrates visualizations of discussion participation, social interaction, assignment progress, and idea trajectory within a unified dashboard for both instructors and students. A key contribution is its support for collective metacognition and socially shared regulation through the Idea Trajectory feature, which represents posts as nodes in an evolving knowledge network to make the joint problem space visible and support monitoring of progress, identification of gaps, and strategic contributions. LLM-generated summaries enhance interpretability and visualization of big ideas, while the Promising Idea function allows participants to highlight contributions with high epistemic potential, directing collective attention and sustaining inquiry. At this stage, CLAto emphasizes making interactional and content-level patterns more visible and actionable; future work will extend its support for explicating epistemic discourses and develop additional AI-powered functionalities. Subsequent research will examine how users interact with and interpret the visualizations, and whether and how the tool improves discourse quality and supports meaningful knowledge construction.

### References (selected)

- Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, 10(4), 345-366.
- Chen, B., & Lin, F. (2020). Representational affordances for collaborative learning in technology-enhanced environments. In *Handbook of learning from multiple representations and perspectives* (pp. 513-531). Routledge.
- Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, 11(3), 263-280.
- Kaliisa, R., Rienties, B., Mørch, A. I., & Kluge, A. (2022). Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. *Computers and Education Open*, 3.
- Lin, F., Chan, C. K. K., & van Aalst, J. (2014). Promoting 5th graders' views of science and scientific inquiry in an epistemic-enriched knowledge-building environment. 11th International Conference of the Learning Sciences, Boulder, Colorado, USA.
- 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.
- Ouyang, F., Zhang, L., Wu, M., & Jiao, P. (2024). Empowering collaborative knowledge construction through the implementation of a collaborative argument map tool. *The Internet and Higher Education*, 62.
- Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Clarendon Press.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). Cambridge University Press.
- Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, 15(3), 3-26.
- Stahl, G., Koschmann, T., & Suthers, D. (2014). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 479-500). Cambridge University Press.
- Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning analytics for online discussion. *Journal of Learning Analytics*, 1(2), 48-71.
- Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). *Learning analytics for online discussions* Proceedings of the Third International Conference on Learning Analytics and Knowledge,