GA2Graph: A Data-Driven Approach to Visualizing and Analyzing Collaborative Learning

Abdussalam Alawini*, Isa Hajara-Yasmin*, Jiabao Xu*, Yutong Zhang† Zhijun Zhao*
*University of Illinois Urbana-Champaign
{alawini, hisa2, jiabaox3, zhijunz3}@illinois.edu
†Stanford University
yutongz7@stanford.edu

***Abstract*—This innovative practice full paper describes an approach to visualizing and analyzing collaborative learning. Collaborative learning is vital for enriching computer science education and developing key skills. However, educators often lack tools for effectively tracking and analyzing student engagement in group work. Our study introduces GA2Graph, a data-driven system that uses Neo4j to visualize and analyze online collaborative patterns in classroom activities. It graphically represents students and their interactions, helping educators understand group dynamics. We tested this system in a large-enrollment database course, revealing that despite emphasis on assigned roles like those in POGIL, students frequently disregarded them and adopted more individualistic or varied collaborative strategies.**

***Index Terms*—Collaborative Learning, Visualization, Educational Data Mining,**

I. Introduction

Collaborative learning (CL) has become a widely discussed and implemented approach in contemporary education. Based on Annett's definition, CL is an education approach to teaching and learning that involves groups of learners working together to solve a problem, complete a task, or create a product [1]. Previous studies emphasize various benefits of collaborative learning, including enhancing problem-solving abilities, fostering a deep comprehension of course material, and nurturing essential interpersonal skills for a successful career in computer science. [1] Computer-Supported Collaborative Learning (CSCL), grounded in the Social Constructivism Theory [2], leverages technologies to facilitate and encourage interactions among students across domains [3]. CSCL has been incorporated into education by various studies [4]–[7]. Previously, there were a lot of studies that emphasized the benefits of CSCL for students, including promoting cognitive strategies, logical thinking, and engagement through social interactions [8]. However, there is limited exploration from instructor aspects, teachers and policymakers may lack understanding of how group collaboration can be effectively integrated into instructional strategies [9].

There a critical problem in which, instructors often lack the tools necessary to monitor and analyze how students collaborate, particularly in relation to predefined roles like those in POGIL. Without this insight, it is challenging to assess the effectiveness of collaborative strategies and ensure that students are engaging in meaningful teamwork. In this project, we aim to explore the support of CSCL for instructors to better understand students' collaboration behaviors and interactions to help them improve their instruction strategy and collaboration policies.

To address this gap, our study introduces an innovative, data-driven system that provides educators with a comprehensive platform for visualizing, analyzing, and interpreting student online collaboration patterns during collaborative learning activities. We utilized the Neo4j Database to conduct a visualization dashboard to help instructors to better track and optimize group work in computer science education. We further investigate our approach in a large-enrollment database course. The course offers various collaborative activities related to database programming and concepts through an online assessment system, PrairieLearn [10]. The interface also implements the group features while encouraging students to use the POGIL [11] roles of Recorder, Manager, and Reflector (see Section 2.2). Based on the graph-based analysis of the log data, we aim to address the following research questions:

RQ1: *How effectively are teams adhering to their self-assigned POGIL roles during group activities?*

RQ2: *What patterns of collaborative learning can be extracted from the GA2Graph graph database to better understand the dynamics and effectiveness of team strategies?*

The contribution of this study is twofold: First, it introduces a dynamic visualization interface that enables instructors to actively explore the collaborative dynamics of their students during in-class group activities, placing them directly in control of the analysis process. Second, it leverages the powerful capabilities of the GA2Graph's Neo4j graph database to examine teams' adherence to POGIL roles and to uncover patterns of collaboration, thereby enhancing the understanding of team strategies.

In the remainder of this article, we will provide an overview of the related work, as well as an introduction to the POGIL roles in Section. II. Section. III demonstrates the workflow and functionality of our tool. The results of quantitative analyses conducted on the course data and the subgraphs generated by our tool are provided in Section. IV, along with discussions on future work and possibilities around this topic. Finally, in Section. V, we conclude this research.

II. Background

A. Related Work

Collaborative learning (CL) is a pedagogical approach where small groups of learners work together to solve a problem, complete a task, or create a product [12]. This approach is designed to capitalize on each participant's resources and skills, asking them to share information, debate concepts, and collectively develop solutions. [12] defines CL as an educational strategy that not only enhances problem-solving capabilities and deepens understanding of the material but also fosters crucial interpersonal skills, which are essential for a successful career in disciplines like computer science.

Building on the principles of CL, Computer-Supported Collaborative Learning (CSCL) integrates technological tools to facilitate and enhance interaction among students [13]. Grounded in Social Constructivism Theory [2], CSCL leverages various technologies to promote communication and collaboration across different learning settings. Studies have shown that CSCL can significantly enhance learning outcomes by promoting cognitive strategies, encouraging logical thinking, and boosting engagement through rich social interactions [14].

Current Works in CSCL Interfaces and Data-Driven Approaches. Recent advancements in CSCL have seen the development of sophisticated interfaces that support data-driven approaches to analyzing collaborative learning. In [14] by Fong et al., they proposed pedagogies for synchronous collaborative learning activities, suitable for both online and in-person classes. For instance, systems employing graph-based modeling techniques have been developed to map student interactions and visualize learning networks, thereby helping educators understand and optimize learning dynamics [15]. CSCL has increasingly incorporated Learning Analytics (LA) [16] and Social Learning Analytics (SLA) [17] to monitor and analyze student interactions. LA and SLA offer valuable insights into how students engage in collaborative activities, using data-driven approaches to identify patterns and optimize learning environments. For instance, Zheng et al. (2021) emphasize the importance of self-regulation and emotion in student interactions with learning analytics dashboards, demonstrating how these tools can enhance the learning experience [18].

In line with the principles of LA, our study contributes to the burgeoning field of educational data mining by introducing GA2Graph. By leveraging data-driven approaches and visualization techniques, our research aims to explore the dynamics of student collaboration, shed light on prevalent interaction patterns, and uncover insights crucial for optimizing collaborative learning experiences. Through our investigation, we seek to bridge the gap between theoretical frameworks like CL and practical implementations, offering educators a powerful tool to understand and enhance collaborative learning environments in disciplines such as computer science with the use of intuitive interfaces.

Graph-Based Modeling in the Computational Social Science Community. The computational social science community has extensively utilized graph-based modeling approaches to study various social networks [19]–[21]. These models are crucial for understanding the complex and dynamic interactions that characterize social systems. In educational settings, these approaches have been adapted to study how students form networks and how these networks influence learning outcomes [22].

This type of modeling provides a robust framework for analyzing the data-rich environments of CSCL systems [23]. By applying similar methodologies, our study aims to contribute to the understanding of how student collaboration can be visualized and optimized through data-driven insights. Furthermore, our use of Neo4J to create a visualization dashboard for instructors highlights the practical application of these complex theoretical models in everyday educational practices. Our research contributes to the field of CSCL by providing a holistic tool that combines advanced data analytics with practical, user-friendly interfaces designed for educators in computer science that utilizes graph-based infrastructure to inspect and analyze collaborative working patterns. This study not only enhances the theoretical foundations of using graph-based models in educational research but also explores the practical implications of these models in real-world teaching environments. The following will elaborate on the contributions of social network analysis using graph-based models, connecting these methodologies to research on collaborative learning in computer science education.

Contributions of Computational Social Science to Graph-Based Social Network Analysis. One of the primary applications of graph-based modeling in computational social science is the analysis of network structures [24]. Researchers use graph theory to quantify relationships between entities (such as individuals or groups) [19], exploring properties like centrality, connectivity, and clustering [19]. For example, studies often measure degree centrality to identify influential individuals within social networks [25]. This analysis helps in understanding how information flows within a group and how social structures can influence behaviors and outcomes.

Graph-based models are also crucial for studying the dynamics of social interactions over time [19]. Dynamic network analysis allows researchers to observe how relationships evolve, how networks react to external changes, and how these changes affect individual and group behaviors. In educational settings, such dynamic analyses can reveal how students interact in small discussion groups [23], as in this case study where Chai et al. explores the dynamics of discussions by tracking the turns in which students talk in small groups.

Building upon the contributions of computational social science to graph-based social network analysis, our research extends this framework into the realm of educational data mining. By applying graph theory and dynamic network analysis techniques to study student collaboration patterns, our work provides a novel perspective on understanding the intricate dynamics of collaborative learning within educational settings. Through the utilization of GA2Graph, our study not only quantifies the relationships between students but also explores how these relationships can evolve over time during in-class group activities.

Visualization of Complex Networks. Visualization is a key aspect of graph-based social network analysis, providing intuitive representations of otherwise abstract data [26]. This is also evident in LA [27]. These visualizations can illustrate complex interdependencies and patterns that might not be apparent from raw data alone. In educational contexts, such visual tools help instructors and administrators visually identify key actors, understand group dynamics, and tailor their teaching strategies to address specific network characteristics. For example, [28] showed how they visualize the socio-gram of the interaction network in different views for easy access. In our study, we aim to prove a similar concept, in which we propose an intuitive interface for student collaboration analysis.

B. Collaborative Assessments in PrairieLearn

PrairieLearn [10] is an innovative assessment system that has been specifically designed to enhance the learning experience through dynamic, adaptive testing and feedback. PrairieLearn [10] supports a wide range of question types, including numeric, graphical, symbolic, programming, and drawing problems. It stands out for its ability to generate a diverse array of parameterized questions, thereby enabling personalized learning paths for students.

A notable feature of PrairieLearn [10] is its capability of facilitating collaborative learning through group assessments. This system allows students to engage deeply with course content alongside their peers, fostering both individual and collective understanding. The process of forming groups in PrairieLearn [10] is straightforward and user-friendly. A student designated to initiate the group can create a new group within the platform and generate a unique join code. This code is then shared with other students, who can use it to join the group seamlessly. This system helps aid in the smooth integration of the POGIL structure.

The POGIL approach is a student-centered teaching method that can employ structured roles to facilitate active learning and collaboration among students [29] and has been shown to have significant success in aiding students to pass course evaluations [11]. In a POGIL classroom, each student in a group is assigned a specific role that comes with its own set of responsibilities, which are critical to the group's success. The roles are essential to our design because they structure how students interact and collaborate, ensuring that each group member contributes meaningfully to the learning process. However, the implementation of roles in CL is not without challenges. One significant issue is ensuring that students adhere to their assigned roles. In practice, some students may dominate the group discussions while others may contribute less, leading to imbalances in participation. Additionally, the effectiveness of roles can vary depending on the group's composition and the complexity of the task at hand. For example, a poorly managed group might struggle to stay on task, or a group without a strong Reflector might fail to critically evaluate their learning process, leading to suboptimal outcomes.

Here are the defined roles used in PrairieLearn [10] for group activities:

Manager: This role is key to organizing the group's activities. The Manager is responsible for ensuring that the group stays on task, meets deadlines, and adheres to the instructions. For example, if a team member is silent for a while, it is the manager's role to check in on the team member. The manager is also responsible for indicating who took on each role at the end of group work. They also act as the primary communicator between the instructor and the group, relaying any important information or questions.

Recorder: The recorder is tasked with documenting the group's answers and insights, the Recorder ensures that all contributions are noted and that the final responses submitted reflect the group's collective understanding. This role is crucial for maintaining a clear and organized record of the group's and they are responsible for being the main "driver," entering answers into PrairieLearn [10]. For example, we encourage the recorder to have PrairieLearn [10] open and share their screen with other team members.

Reflector: The Reflector's role involves thinking about the group's learning process. They consider what strategies are working, what challenges the group is facing, and how the group's dynamics are affecting the learning outcomes. After each session, the Reflector provides feedback in a short survey that describes the teams performance which can be used to improve future group interactions. The Reflector is also responsible for ensuring that all team members understand what's going on. For example, the reflector might ask whether each team member understands the problem before moving on to the next problem.

Contributor: Every member of the group, regardless of their primary role, is expected to contribute ideas, ask questions, and provide solutions. Contributors are active participants, ensuring a rich, diverse discussion and broad exploration of the material. Contributors are members with no team organization duties but are expected to contribute to the teamwork. They are also tasked with providing peer reviews periodically during the semester.

Before starting an activity, each group member selects their role from the available options. This selection is typically done through a simple interface on PrairieLearn [10] where roles are chosen, and students can see which roles have already been filled. While the system facilitates the assignment of roles, it does not enforce them rigidly nor do the instructors. For example, while the Recorder is generally the one to submit the group's answers, other roles like the Manager might occasionally take on this task if needed. This flexibility allows groups to adapt to the dynamics of their specific members and situations, maintaining fluidity in their collaborative process.

Each student within a group can submit solutions to the problems presented. Although all group members are working on the same set of problems, they do not necessarily see the same view on their screens. Students do not edit the same solution synchronously; they can only view their teammates' submissions in the submission log. This setup ensures that each student can track the group's progress and contributions, fostering a sense of collective effort and responsibility. This system allows for a continuous exchange of ideas and strategies, as students can learn from the attempts of others. After a submission, immediate feedback is provided. If the answer is correct, the problem is marked as correct for the entire group.

However, PrairieLearn [10] allows students to continue attempting the problem even after achieving the correct answer. This feature is particularly beneficial for exploratory learning, where students can test different approaches or solutions without the risk of affecting their grade. The process fosters a dynamic interaction among students, as they can discuss strategies, correct misconceptions, and collectively seek better understanding or alternative solutions.

III. System Overview

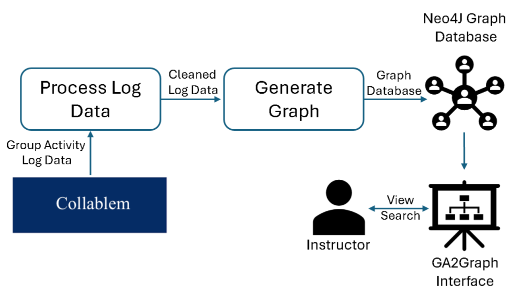


Fig. 1: GA2Graph Architecture: PrairieLearn's [10] group activities data is processed and transformed to a graph format and stored in Neo4j Graph Database. The GA2Graph GUI connects with the Neo4j backend and enables users to visualize and filter the graph.

Figure 1 illustrates the architecture of the GA2Graph system. Data from PrairieLearn's [10] (PL) group activity logs is fed into the GA2Graph system, where it is processed and transformed into a graph format. Users can dynamically interact with this graph to examine collaboration patterns and conduct advanced searches for specific patterns of interest. Clicking on any node or edge in the graph provides a detailed (zoomed-in) view of a particular group. The design and architecture of the graph database are discussed in the following section.

A. Processing PrairieLearn Log Data

PrairieLearn [10] captures a detailed log of students' activities for collaborative assignments. Each log entry includes the student identifier, time of submission, the attempted question, and the grading result. This grading information is stored in a JSON document that details whether the answer was correct or incorrect, along with any feedback provided, such as error messages from the relational database system for SQL queries. Additionally, the record includes an event attribute describing the log event. Notable events include "Group Create" when a student forms a new group, "Group Join" when a student joins an existing group, "View Variant" when a student views a variant of a collaborative assignment question, and "Submission" when a solution is submitted to the question.

Process Log Data subsystem extracts data from PrairieLearn's [10] group activity logs (CSV files), capturing each team's group interactions. Initially, it filters out all events except for "Submitted" events, as these contain the students' solutions and the results of the grading. For each relevant log entry, it parses the JSON grading result document to extract 1) the submission score, 2) the grader feedback, and 3) the submitted solution. The processed log data is then forwarded to the "Generate Graph" process, which is described in the subsequent section.

B. Generating a Graph Database

The **Generate Graph** subsystem utilizes the processed log data to create a graph database. The outcome of the PrairieLearn [10] to Graph subsystem is a comprehensive graph encapsulating all team interactions within a particular group activity. This graph consists of several subgraphs, each representing a different team. Within these subgraphs, each node represents a student, and each edge denotes a submission to a GA question. This overview provides instructors with a bird's-eye view (see Figure 2) of the entire class on a single canvas, enabling a clear visual assessment of collaborative dynamics.

For every team, the **Generate Graph** subsystem creates a graph node for each member involved in the activity and adds information related to their collaborative activities. This information includes: student identifier; the group name; the role of this student; an array of the questions this student attempted; all the questions that were first answered correctly by this student. Roles and groups are stored in two arrays in a node to avoid creation of separate nodes for the same student. This not only reduces the data size, but also establishes a one-to-one mapping between the students and their nodal instances in the database, which is a logically sound approach. Indexing would be use when extracting the role or group information of a student in a specific group assignment.

Graph relationships (edges) are established based on question submission activity. The **Generate Graph** subsystem generates a directed edge from node A to node B if student A submits an answer to the same question after student B. If a student submits two consecutive answers to the same GA question, a self-directed edge (i.e., self-loops or self-edges are edges from a node A to itself) is created. If an edge already exists between two nodes, the subsystem does not create a new one; instead, it increases the collaboration count property on the existing edge. Additional properties presented on the edge include: the identifier of the student who submitted the current answer; the identifier of the student who submitted the previous answer; the group name to which both students belong; the question id; the number of the group assignment during which this submission to a question occurred; an array indicating correct/incorrect results of previous submissions to this question; and an array of all time intervals between two consecutive submissions to this question from the same pair of students.

C. GA2Graph Interface

GA2Graph Interface allows users to observe and filter the directed graph produced by the **Generate Graph** subsystem. GA2Graph Interface consists of a scalable graphical interface and an advanced graph search panel.

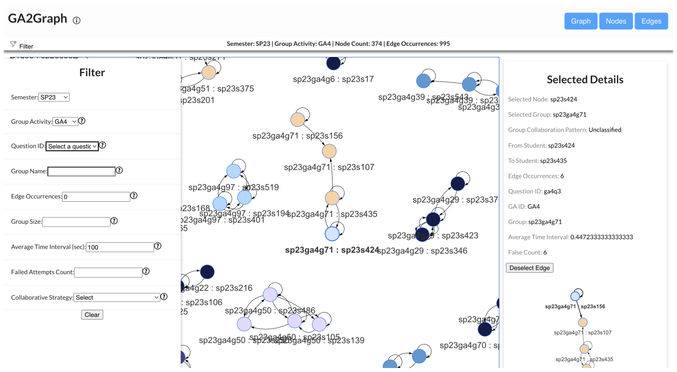


Fig. 2: GA2Graph interface: the GUI allows users to select a specific semester and GA on the top filter bar and displays the graph in the main screen. On the right side is the detailed information of the selected node and edge

The graphical interface exhibits a directed collaboration graph corresponding to a designated group assignment within a specified academic term. In this representation, each node denotes an individual student, while the edges shows temporal collaborative relationship between them. Graph nodes are color-coded based on group to highlight the different collaboration patterns among groups. Graph edges are designed to be non-overlapping, allowing a clear visualization on bidirectional edges. As shown in Figure 2, when a user clicks on a node or an edge, an additional panel appears on the right side of the interface. This panel displays general group information such as "Question ID", "Group Name", "Edge Occurrences", and etc on the left, along with a subgraph that only includes the nodes and edges of the selected subgraph (team).

The left filter panel (Figure 3) has two major functions: graph selection and graph filter. Users can filter the graph by variables like "Semester" and "GA" (group assignment). Upon selection, the graph window will update correspondingly. For advanced analysis, "Question ID" allows users to observe the collaboration graph on a selected question; "Group Name" allows users to only observe the behavior or a specified group; "Edge Occurrences" represents the frequency of the collaboration; "Group Size" filters the groups by the number of students in; "Average Time Interval" filters the time interval among all the submission between a pair of students in seconds; "Failed Attempts Count" filters the number of tries the student took before got the question correct. "Collaborative Strategy" filters the strategy used by the group

To demonstrate a practical application, In Figure 3, we choose the directed graph that corresponds to spring2023-ga4. Upon selection, the graph reveals a high level of intra-group connectivity among members. Given the diversity in question types, our focus shifts to examining collaboration specifically related to a SQL coding question (question 2) with groups of size 3 having less than or equal to 3 wrong tries and 20 seconds submission interval. In Figure 4, we show the updated graph accordingly. We observed a drop of students count from 374 to 31 and majority of the groups retained tend to collaborate within groups rather than having one student submit all the attempts.

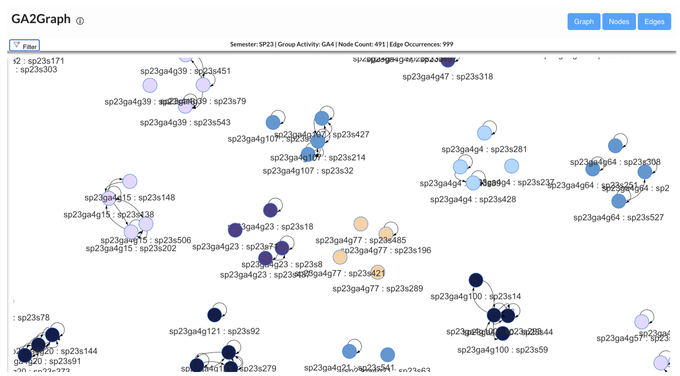


Fig. 3: Part of the GA graph of semester 2023-GA4

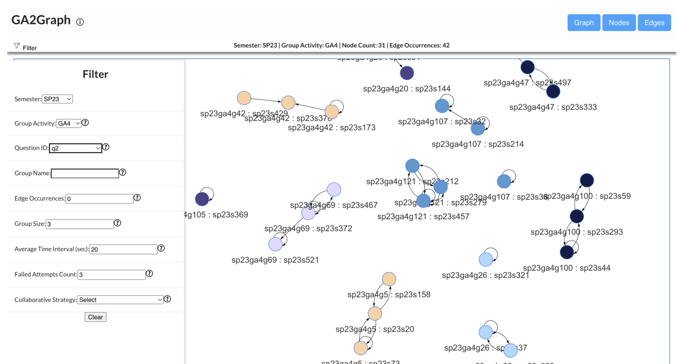


Fig. 4: The GA graph of semester 2023-GA4 on question2, with Group Size 3, Failed Attempts Count 3, and Average Time Interval (sec) 20

IV. Exploring Collaborative Behaviors with GA2Graph

In this section, we utilize the graph database from the GA2Graph system to examine collaborative learning patterns and behaviors. We begin by outlining the datasets employed for this analysis. Next, we assess how students adhered to the POGIL rules during their collaborative activities. Finally, we investigate basic collaborative learning patterns extracted from the GA2Graph database.

A. Data Collection

Our data is collected from the Database Systems course offered to upper-level undergraduate and graduate Computer Science students at the University of Illinois Urbana-Champaign. The data spans from the Fall 2022 semester, with student enrollment at 514, and the Spring 2023 semester, with enrollment at 470. Instruction was delivered in person using a flipped-classroom model. Pre-recorded lecture videos, including a brief knowledge check quiz, were assigned for students to review before class. During class, the instructor reviewed the quiz solutions, worked through practice problems, and addressed student questions or misunderstandings. The remainder of the class time was dedicated to collaborative group exercises, allowing students to deepen their understanding of the concepts presented in the pre-recorded lectures.

The in-class group exercises were labeled as ”GA” (Group Activity). To maintain the course consistency, the GA material remained unchanged in the Fall 2022 and Spring 2023 semester. The GA activities were structured into seven groups, a total of 16 assignments, which align with the course timeline. These groups were categorized as follows: SQL (5 assignments), Database Design (2 assignments), Indexing and Storage (1 assignment), Transaction Management (1 assignment), Query Processing and Optimization (2 assignments), MongoDB (2 assignments), and Neo4j (2 assignments). Each GA comprised a set of fixed questions and questions tailored to the specific material. The fixed questions included questions such as role distribution, activity reflection, and peer evaluation, all designated for specific roles such as manager or reflector. The material-tailored questions test a wide range of topics based on the course content. These questions include autogenerated coding exercises in SQL, MongoDB, and Neo4j, as well as drawing questions, multiple-choice queries, and autogenerated, autograded short-answer assessments.

B. POGIL Roles Analysis

As mentioned in Section II-B, the course that provided our dataset utilizes POGIL roles without strict enforcement. These roles were implemented to structure team collaborations. Therefore, we aim to examine how students adhered to these roles, as this is crucial for understanding the work dynamics within each team.

Figure 5 provides an analysis of the percentage of teams adhering to POGIL roles across 15 Group Activities (GAs) during the Fall 2022 (FA22) and Spring 2023 (SP23) semesters. From the chart, it is evident that in SP23, adherence to POGIL roles started relatively high, with around a quarter of the teams following the roles during the first GA. However, there is a sharp decline in compliance with POGIL roles as the semester progresses, with the percentage dropping quickly and stabilizing around the 5% mark by GA 5. The adherence then fluctuates slightly but remains generally low throughout the rest of the semester, with an increasing trend after GA 12.

In contrast, FA22 starts at a lower compliance level, slightly below 16%, and also experiences a decline, though less steep than in SP23. The adherence levels out at around 6% by GA 4, showing a somewhat consistent yet fluctuates slightly through subsequent GAs. Towards the end of the semester, the percentage of teams following the POGIL roles in SP23 shows a more noticeable increase compared to the FA22 semester.

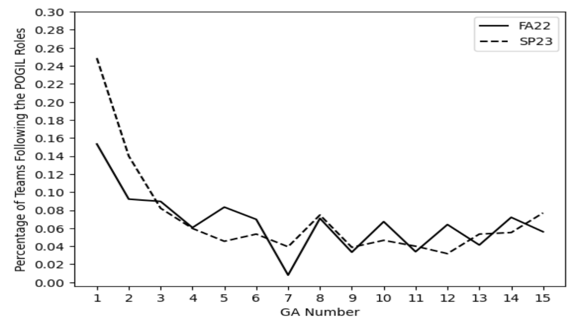


Fig. 5: Role adherence line graph in Fall 2022 and Spring 2023 semester

The overall trend suggests a significant drop-off in the adherence to POGIL roles after the first few group activities in both semesters, with a more stable but lower adherence in FA22 compared to SP23. This pattern may indicate initial enthusiasm or compliance with structured roles that diminishes over time, possibly due to the natural dynamics of team development or the perceived relevance or effectiveness of these roles in collaborative tasks. The slight increase towards the end of SP23 could suggest a recalibration of team strategies or a renewed focus on the roles' benefits as students prepare for final assessments or projects. This data highlights potential areas for educational interventions to sustain or increase the effectiveness of POGIL roles in team-based learning environments.

The results from the graph clearly indicate that most teams did not adhere to the assigned POGIL roles, which provides a strong motivation for our analysis of collaborative patterns. This lack of adherence likely contributes to a diversity of team behaviors, potentially affecting the overall effectiveness and dynamics of group activities. Understanding these varied behaviors is crucial as they reflect the direct impact of non-compliance with structured roles on team performance and interactions.

C. Collaborative Patterns Analysis

Motivated by our analysis of adherence to POGIL roles, we now shift our focus to examining the collaborative learning patterns exhibited by teams across 15 GAs during the FA22 and SP23 semesters. Our study is further informed by instructor observations of team behaviors during these group activities. We have identified four primary types of collaborative patterns for this analysis:

- **Single Member Contributor**: In this pattern, only one student submits solutions to all the GA questions, potentially indicating that the team might be following the POGIL roles. However, this pattern could also include scenarios where one student predominantly takes the lead,

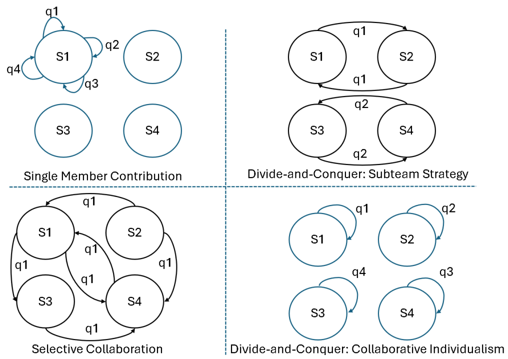


Fig. 6: Classification of Team Strategies for Group Activities

regardless of whether the team formally adheres to the POGIL roles.

- **Divide-and-Conquer: Subteam Strategy**: Here, teams split into smaller subgroups, with each subgroup tackling a different problem within the GA. This strategy facilitates focused problem-solving within subteams.
- **Divide-and-Conquer: Collaborative Individualism**: This pattern involves each team member working independently on separate problems of the GA. It highlights individual contribution within a collective framework.
- **Selective Collaboration**: Teams displaying this pattern engage collaboratively on at least one problem of the GA, showing selective but significant cooperative effort.

These categories help us understand the diverse approaches teams take in collaborative environments, potentially influencing the overall effectiveness and dynamics of group activities.

Figure 7 illustrates the distribution of various collaborative strategies across 15 Group Activities (GAs) during the FA22 (Fig. 7-a) and SP23 semesters (Fig. 7-b). Each chart categorizes the percentage of teams adopting one of four collaborative patterns: *Single Member Contribution*, *Selective Collaboration*, *Divide and Conquer: Collaborative Individualism*, and *Divide and Conquer: Subteam Strategy*. We now discuss key observations from the charts.

In Fall 2022 (Fig. 7-a), *Single Member Contribution* strategy exhibited moderate yet decreasing involvement, peaking early in the semester at about 17% and demonstrating a general downward trend. On the other hand, *Divide and Conquer: Collaborative Individualism* was prominently adopted, particularly in the middle of the semester, where it reached its highest at around 60%, suggesting a strong preference for this collaborative approach as group tasks potentially grew more complex. *Selective Collaboration* showed notable peaks early and towards the end, with the highest point being around 50%, indicating periods of increased collaborative interaction. *The Divide and Conquer Subteams* strategy was the least adopted, maintaining a consistently low presence around 6-8%, suggesting limited effectiveness or appeal.

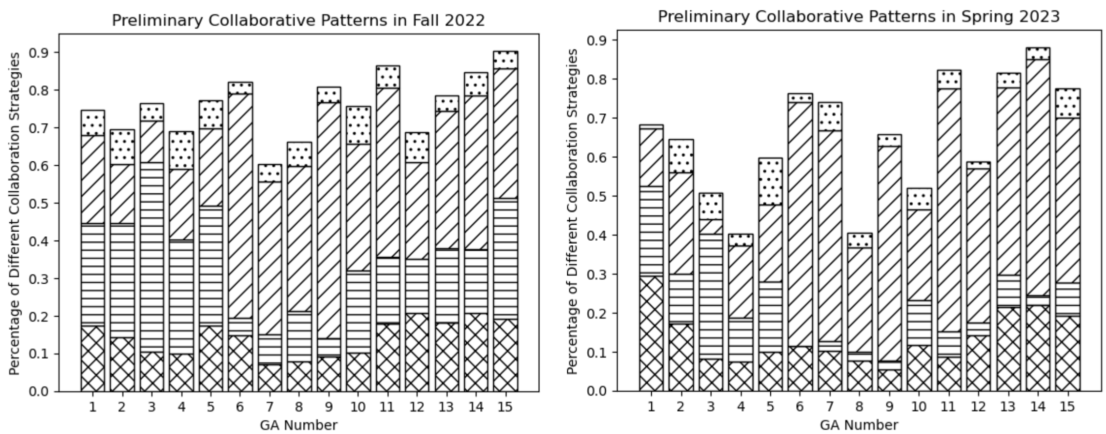
During Spring 2023 (Fig. 7-b) , *Single Member Contribution* started high at around 29% and showed more stability, suggesting a consistent reliance on individual contributions. *Divide and Conquer: Collaborative Individualism* remained the most popular strategy, especially mid-semester, peaking at 62%, reinforcing its favorability among students for handling complex tasks. In contrast, *Selective Collaboration* observed a significant decrease, particularly noticeable with zero adoption in several GAs, indicating a shift away from less structured collaborative forms. *Divide and Conquer Subteams* continued to exhibit minimal use, echoing the trend from Fall 2022 with very low percentages throughout.

Divide and Conquer: Collaborative Individualism dominated in both semesters, suggesting it as the preferred method for managing collaborative tasks, possibly due to its balance between individual responsibility and group interaction. On the contrary, *Selective Collaboration* faced the most significant fluctuation, with a marked decrease in Spring 2023, which could imply a shift in student preferences or effectiveness of the strategy over time. *Single Member Contribution* and *Divide and Conquer Subteams* displayed opposite trends, with the former showing relatively high and stable adoption and the latter remaining consistently low.

The data from both semesters indicates a strong preference for strategies that allow for individual accountability within a collaborative framework, as seen with *Divide and Conquer: Collaborative Individualism*. The consistent low interest in *Divide and Conquer Subteams* suggests that while teams value structured teamwork, they prefer strategies that also allow individual contributions to stand out. The shift in *Selective Collaboration* usage underscores the dynamic nature of team collaboration, emphasizing the need for educational strategies that adapt to evolving student preferences and project demands. These insights could guide educators in designing group activities that better align with observed student behaviors and preferences for effective teamwork.

Collaborative Patterns in Relation to GA Question Types. The observed strategies in GAs 1 to 5, where SQL coding problems allowed unlimited attempts, show a clear tendency in both semesters for teams to favor Single Member Contribution and Divide and Conquer: Collaborative Individualism. This indicates that the freedom to make unlimited attempts without team penalties encouraged either strong individual contributions or individual efforts within a broader team framework. These strategies were particularly prominent in both Fall 2022 and Spring 2023, reflecting a consistent approach to tasks that allow for experimental problem-solving.

For GAs 9 to 11, which featured autogenerated and autograded questions requiring teams to start over with a new variant upon making mistakes, there was a notable increase in Selective Collaboration in both semesters. This shift underscores the strategic adaptation of teams to engage in more integrated collaborative efforts to avoid the costs associated with errors. The reduction in Single Member Contribution during these periods further illustrates a move towards collective problem-solving in response to the task's constraints. Lastly, in GAs 12 to 15, with the introduction of complex database coding tasks involving MongoDB and Neo4j, there was a resurgence in Divide and Conquer: Collaborative Individualism in both



(a) Fall 2022: Single Member Contribution peaked early at 17%, showing a downward trend. Divide and Conquer: Collaborative Individualism peaked mid-semester at 60%. Selective Collaboration peaked early and late at 50%, while Divide and Conquer Subteams remained low at 6-8%

(b) Spring 2023: Single Member Contribution started high at 29%, remaining stable. Divide and Conquer: Collaborative Individualism peaked mid-semester at 62%. Selective Collaboration decreased significantly, often hitting zero. Divide and Conquer Subteams remained low throughout.

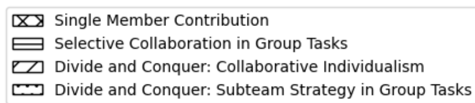


Fig. 7: Collaboration strategy decisions among 15 GAs in Fall 2022 and Spring 2023

semesters, suggesting a return to strategies that allow for specialization and individual expertise within complex and technical contexts. This pattern reaffirms that the nature of the tasks significantly influences the choice of collaborative strategies, consistently across both observed semesters.

V. Limitation and Future Work

A. Limitations

Our study presents a novel approach to analyzing student collaboration patterns using submission timestamps as proxies for interactions. However, this method has several limitations:

Reliance on Submission Timestamps. We primarily use submission timestamps to identify collaborations between students. While this provides a temporal view of interaction, it does not capture the content or quality of the collaboration. Future work should include a more detailed analysis of the submitted solutions to determine whether students are building on each other's work or mere submitting simultaneously. For instance, by comparing code similarities and changes over time, we can gain deeper insights into the collaborative process.

Generalizability of Findings. Our research is based on data from the university of Illinois, known for its strong Computer Science program. This specific context of may limit the generalizability of our findings. Students at different universities or in different programs may exhibit different collaboration patterns due to varying curricula, teaching methods, and institutional cultures. Therefore, it is essential to replicate this study in diverse educational settings to validate our results and ensure broader application. To enhance the generalizability of our findings, future research should collect and analyze data from a variety of educational institutions, including those with different sizes, locations, and learning content. Comparing collaboration patterns across these diverse contexts will help identify universal principles of effective collaboration as well as context-specific strategies.

Focus on Specific Interaction Metrics. Our current approach focuses on submission times without considering other possible interaction metrics such as communication logs, forum posts, or peer evaluations. Incorporating these additional data sources could provides a more comprehensive understanding of student collaboration dynamics.

B. Future Work

To address the limitation of relying solely on submission timestamps, future work should integrate code analysis techniques. Leveraging methods from [30] such as Levenshtein Edit Distance, we can analyze code changes to identify instances where students are truly collaborating by building on each other's solutions. This could involve detecting patterns of code reuse, modification, and joint problem-solving efforts, as done by Anonymous 2021. Additionally, it may also be beneficial to analyze and present how the evolution of a collaborative group may have contributed to the final submitted results [31]. As investigated by Anonymous 2022, tracing assignment submissions can provide insight into how students learn, and the same idea can be applied to study how different collaboration patterns may affect learning patterns.

It is crucial to evaluate the effectiveness and usability of our system, GA2Graph, in real-world educational settings.

Conducting user studies with instructors can help assess how well the system aids in identifying collaborative learning patterns and informs instructional strategies. Feedback from these studies will be invaluable for refining the system to better meet educators' needs.

To further support collaborative learning, we propose developing a dashboard for students that allows them to monitor their collaboration activities. This dashboard could provide real-time feedback and analytics on their participation and collaboration patterns, encouraging more effective teamwork and self-regulation. Features could include visualizations of collaboration networks, alerts for isolated students, and recommendations for improving collaborative efforts.

Beyond submission timestamps, future studies should explore incorporating other interaction metrics such as communication logs from discussion forums, chat histories, and peer feedback mechanisms. These additional data points can provide a richer and more nuanced understanding of how students collaborate and interact in online and hybrid learning environments.

VI. Conclusion

We developed GA2Graph, an interactive system that visualizes student collaborations from online assessments to identify group patterns and adherence to POGIL roles. Our preliminary results show that despite the emphasis on POGIL roles, students frequently disregarded them, preferring independent or alternative collaborative strategies. Notably, the 'Divide and Conquer: Collaborative Individualism' strategy was preferred, indicating a trend towards balancing individual accountability with teamwork. These findings suggest that POGIL roles have minimal impact on actual collaboration

References

[1] N. Annett, "Collaborative learning: definitions, benefits, applications and dangers in the writing center," *University of Richmond, Virginia, USA. Retrieved Oct*, vol. 5, p. 2012, 1997.

[2] P. Adams, "Exploring social constructivism: Theories and practicalities," *Education*, vol. 34, no. 3, pp. 243–257, 2006.

[3] H. Jeong, C. E. Hmelo-Silver, and K. Jo, "Ten years of computer-supported collaborative learning: A meta-analysis of cscl in stem education during 2005–2014," *Educational research review*, vol. 28, p. 100284, 2019.

[4] G. Herman, Y. Jiang, Y. Jiang, S. Poulsen, M. West, and M. Silva, "An analytic comparison of student-scheduled and instructor-scheduled collaborative learning in online contexts," in *2022 ASEE Annual Conference & Exposition*, 2022.

[5] Y.-T. Lin, C.-C. Wu, and C.-F. Chiu, "The use of wiki in teaching programming: Effects upon achievement, attitudes, and collaborative programming behaviors," *International Journal of Distance Education Technologies (IJDET)*, vol. 16, no. 3, pp. 18–45, 2018.

[6] T. Faas, L. Dombrowski, A. Young, and A. D. Miller, "Watch me code: Programming mentorship communities on twitch. tv," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–18, 2018.

[7] J. P. Ucan, O. S. Gomez, and R. A. Aguilar, "Assessment of software defect detection efficiency and cost through an intelligent collaborative virtual environment," *IEEE Latin America Transactions*, vol. 14, no. 7, pp. 3364–3369, 2016.

[8] J. Lämsä, R. Hämäläinen, P. Koskinen, J. Viiri, and E. Lampi, "What do we do when we analyse the temporal aspects of computer-supported collaborative learning? a systematic literature review," *Educational Research Review*, vol. 33, p. 100387, 2021.

[9] G. Stahl, "A decade of cscl," *International Journal of Computer-Supported Collaborative Learning*, vol. 10, pp. 337–344, 2015.

[10] University of Illinois Urbana-Champaign, "Prairielearn," https://prairielearn.engr.illinois.edu/, 2024, accessed: 2024-08-18.

[11] L. Walker and A.-R. M. Warfa, "Process oriented guided inquiry learning (pogil®) marginally effects student achievement measures but substantially increases the odds of passing a course," *PLoS One*, vol. 12, no. 10, p. e0186203, 2017.

[12] B. L. Smith and J. T. MacGregor, "What is collaborative learning," 1992.

[13] E. Lehtinen, K. Hakkarainen, L. Lipponen, M. Rahikainen, and H. Muukkonen, "Computer supported collaborative learning: A review," *The JHGI Giesbers reports on education*, vol. 10, p. 1999, 1999.

[14] M. M. Fong, L. Butler, A. Alawini, G. L. Herman, and M. Silva, "Board 254: Developing tools, pedagogies, and policies for computer-based collaborative learning activities," in *2023 ASEE Annual Conference & Exposition*, 2023.

[15] B. Grawemeyer, A. Wollenschlaeger, S. Gutierrez-Santos, W. Holmes, M. Mavrikis, and A. Poulovassilis, "Using graph-based modelling to explore changes in students' affective states during exploratory learning tasks," 2017.

[16] D. Clow, "An overview of learning analytics," *Teaching in Higher Education*, vol. 18, no. 6, pp. 683–695, 2013.

[17] S. B. Shum and R. Ferguson, "Social learning analytics," *Journal of educational technology & society*, vol. 15, no. 3, pp. 3–26, 2012.

[18] J. Zheng, L. Huang, S. Li, S. P. Lajoie, Y. Chen, and C. E. Hmelo-Silver, "Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard," *Computers & Education*, vol. 161, p. 104061, 2021.

[19] G. Jethava and U. P. Rao, "User behavior-based and graph-based hybrid approach for detection of sybil attack in online social networks," *Computers and Electrical Engineering*, vol. 99, p. 107753, 2022.

[20] X. Li, L. Sun, M. Ling, and Y. Peng, "A survey of graph neural network based recommendation in social networks," *Neurocomputing*, vol. 549, p. 126441, 2023.

[21] S. Bayrakdar, I. Yucedag, M. Simsek, and I. A. Dogru, "Semantic analysis on social networks: A survey," *International Journal of Communication Systems*, vol. 33, no. 11, p. e4424, 2020.

[22] D. Sun, F. Ouyang, Y. Li, and H. Chen, "Three contrasting pairs' collaborative programming processes in china's secondary education," *Journal of Educational Computing Research*, vol. 59, no. 4, pp. 740–762, 2021.

[23] A. Chai, J. P. Le, A. S. Lee, and S. M. Lo, "Applying graph theory to examine the dynamics of student discussions in small-group learning," *CBE—Life Sciences Education*, vol. 18, no. 2, p. ar29, 2019.

[24] J. Zhang, W. Wang, F. Xia, Y.-R. Lin, and H. Tong, "Data-driven computational social science: A survey," *Big Data Research*, vol. 21, p. 100145, 2020.

[25] F. Bloch, M. O. Jackson, and P. Tebaldi, "Centrality measures in networks," *Social Choice and Welfare*, vol. 61, no. 2, pp. 413–453, 2023.

[26] S. Majeed, M. Uzair, U. Qamar, and A. Farooq, "Social network analysis visualization tools: A comparative review," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE, 2020, pp. 1–6.

[27] P. J. Giabbanelli, A. A. Tawfik, and V. K. Gupta, "Learning analytics to support teachers' assessment of problem solving: A novel application for machine learning and graph algorithms," *Utilizing learning analytics to support study success*, pp. 175–199, 2019.

[28] Y. Wang and F. Qiu, "A web-based system for visualizing and analyzing interaction structure in online collaborative learning," in *Advances in Web Based Learning–ICWL 2009: 8th International Conference, Aachen, Germany, August 19-21, 2009. Proceedings 8*. Springer, 2009, pp. 424–433.

[29] R. S. Moog, J. N. Spencer, and A. R. Straumanis, "Process-oriented guided inquiry learning: Pogil and the pogil project," *Metropolitan Universities*, vol. 17, no. 4, pp. 41–52, 2006.

[30] S. Yang, Z. Wei, G. L. Herman, and A. Alawini, "Analyzing patterns in student sql solutions via levenshtein edit distance," in *Proceedings of the Eighth ACM Conference on Learning@ Scale*, 2021, pp. 323–326.

[31] S. Yang, G. L. Herman, and A. Alawini, "Analyzing student sql solutions via hierarchical clustering and sequence alignment scores," in *1st International Workshop on Data Systems Education*, 2022, pp. 10–15.