Modeling Student Collaboration Network to Enhance Student Interactions

1st Hemraj Ojha
Computer Science
Tribhuvan University
Kathmandu, Nepal
hemraj.ojha@kathford.edu.np

2nd Vijayalakshmi Ramasamy
Computer Science
University of Wisconsin-Parkside
Kenosha, Wisconsin, USA
ramasav@uwp.edu

3rd James Kiper
Computer Science
Miami University
Oxford, Ohio, USA
kiperjd@miamioh.edu

4th Gursimran S. Walia
School of Computer and Cyber Sciences
Augusta University
Georgia, USA
gwalia@augusta.edu

***Abstract*—The Covid-19 pandemic, as Henry Kissinger mentions, will not only "forever alter the world order," but also potentially transform the ever-changing higher education world. The recent increase in technological innovations in information, communications, and computer technology has profoundly transformed traditional teaching-learning processes and peer-to-peer interactions for knowledge transfer. One such radical change in technology that researchers are continually working on is motivating collaborative learning and student interactions to improve their learning experiences. Collaborative Learning (CL), where students work in groups to achieve a specific learning objective, can facilitate a deep learning activity that promotes student participation. However, the potential of discussion forums is limited due to their unstructured nature in LMSs like Canvas.**

We propose and develop a structured discussion forum that can offer a platform to communicate and discuss problems and receive feedback, discuss solutions, and suggestions online. Students who participate in these discussion forums can benefit in multiple ways, including increased class preparedness and more active learning. The twofold objectives and outcomes include 1) analyzing discussion board data to reveal students' interaction and their degree of participation in the course, and 2) developing a toolset to draw useful inferences from such collaboration networks. Specifically, our schema-based model can help students visualize the discussion board networks creating an engaged learning environment. Furthermore, the model can help draw valuable inferences of the patterns of student interactions and assess student participation and belonging in the course with greater precision.

This paper demonstrates a schema-based discussion board model that can allow researchers to collect better-formatted discussion data and more reliable information about the posts, such as the type of posts and the relationships of each post with others. The reimagined discussion boards include the ability to classify discussion posts using various parameters, visualize the posts' patterns of interactions, identify their relationships with other discussion posts, and precisely evaluate student participation in discussions to monitor the major topics of discussion. We believe that the result of increased participation in discussions with other students will have the effect of increasing students' sense of belonging to the community of scholars.

***Index Terms*—Collaboration network, complex network analysis, structured collaboration network**

I. Introduction

Collaborative Learning is an approach in which individuals work in groups to achieve a specific objective. In educational/instructional collaborative learning, teachers and students make a combined effort that involves working together to acquire both problem solving skills and additional domain knowledge. Collaborative learning helps to assimilate new knowledge that enhances the productivity of a course and makes it more result-oriented. Collaborative learning research thus involves creating efficient tools and techniques to enable student engagement, developing critical skills of how to think about and evaluate their knowledge. These skills are critically necessary in foundation-level Science, Technology, Engineering, and Mathematics (STEM) courses. One of such tools is a discussion forum. Effective collaborative learning methods have offered a technique for constructive engagement among university students that has many positive effects not only on students' problem solving, critical thinking, social interaction, and perseverance but also on students' academic achievement [1]. A discussion forum is a learning network for the students to communicate and share knowledge about the course content. Students can ask questions about concepts, give examples of how they encounter class material in their everyday lives, or discuss difficult topics to gain greater understanding. Students who participate in discussion forums benefit in multiple ways, including increased class preparedness and more active learning [2]. Therefore, discussion forums not only offer a platform to communicate and discuss problem solving strategies, give and receive feedback, discuss solutions, and provide suggestions online but also help improve belonging in the classroom setting.

In traditional discussion forums, each discussion board contains a collection of threads, each of which contains many postings in a linear order related to the learning experience of the students. However, these threads are not properly

synchronized since all threads are ordered linearly, and the flow of information is not clearly understood. The systematic flow of these discussion threads should help board readers better, understand the flow of conversations, and thereby respond to the relevant posts more effectively. However, in the way typically designed, traditional discussion boards do not provide immediate insight into the interaction patterns of students. Replying or posting to a specific student is also not clearly represented, as reply threads maintain a linear order. Therefore, to analyze and interpret these discussion boards, some additional steps must be taken. Discussion boards can be efficiently modeled as collaborative learning networks using graph mining to gain insights into students' behavior and interactions. Social Network Analysis (SNA) helps visualize interactions as a graph network and study the relationships and cohesion between individuals forming large communities [3]. Collaborative learning strategies result in greater participation of the student in the learning process [4]. Therefore, modeling discussion forums into collaborative networks is essential.

The emergence of new technological advancements including Computer-Supported Collaborative Learning (CSCL) tools has made it possible for instructors to implement and practice productive collaborative learning strategies in the curriculum by devising highly structured goal-oriented discussion forums to encourage students to discuss course concepts and at the same time acquire new skills and enrich knowledge. Designing well-crafted and more structured discussion boards are often the key to establish effective conversations in the context of knowledge sharing. The objective of this work is to propose a structured schema-based discussion board to promote the effective use of class discussion.

The research study aims to efficiently model discussion boards as collaborative learning networks using Social Network Analysis to visualize the interactions as a graph network and to study the relationships and cohesion between individuals who form large communities. Here we present a new model based on our preliminary research [5], which can allow researchers to investigate the new hypothesis. The inferences drawn from the dynamics of student network interactions using unstructured discussion forums are neither beneficial nor practical. There is no systematic or predictable flow of information observed in the participant posts in these types of discussion boards which are often termed as unstructured discussion boards.

It is challenging to model these discussion forums with graph networks. The threads of discussion posts are located one after the other in a linear structure, which makes it difficult to precisely identify the direction (source and destination) of the posts and characterize the posts by response type/purpose (whether it is a comment, question, solving a problem, and others). Therefore, the graph networks that are built from unstructured discussion forums do not represent the flow of communication of discussion posts or depict the specifics of students' interaction patterns. As a result, the inferences drawn will be insufficient. By representing student discussion boards using a structured and schema-based model, it is possible to visualize the discussion board collaboration networks more precisely and draw more reliable and useful inferences on the patterns of student interactions.

Schema-based modeling enables representing the flow of discussion posts in a structured way to facilitate increased participation by making the communication more flexible. It enables an effective user interface that allows participants to choose the type of posts they write and at the same time visualize all posts written by their peers with additional information about who replied to which post and the type of each post, etc. Thus, by using schema-based discussion board models, researchers can collect better formatted discussion data along with more reliable information about the posts, such as the type of posts and the relationships among posts. It becomes easy to classify the posts by using various parameters, visualize the patterns of interactions of the posts, and identify their relationships with other posts. Creating a specific model for discussion boards and applying network-based data analytic techniques to the data collected using schema-based models help to measure student participation and effective information dissemination across the entire network. Thus, our working hypothesis is that a schema-based structured approach to discussions will produce an increase in student participation and consequently improve student engagement in learning. It is this hypothesis that we investigate here.

The rest of the paper is organized as follows. Section II provides an overview of collaborative learning techniques in educational environments, modeling transaction data as graphs and graph clustering techniques in the literature. Section III describes the proposed framework that models the unstructured discussion data as a graph to analyze the characteristics of the network based on measures of similarity. In Section IV, a detailed discussion of the experimental results is presented, including the Discussion Board User Interface Tool and is finally followed by a summary of findings and future directions in Section V.

II. Background

A. Related Work

Various educational researchers in the past have realized the importance of discussion in learning to promote higher-order thinking skills that involve analysis, evaluation, and synthesis [6] [7] [8]. Online discussions offer a method to construct effective course pedagogy by transforming traditional teacher-centered learning into student-centered learning.

Many researchers have studied asynchronous discussion forums. In moderated threads, researchers [9] have observed interesting patterns (cycles) of interactions that indicate active collaborative learning; however, in unmoderated threads, the interactions were more linear rather than having interesting cycles. When studied within controlled environments, some research [10] indicates that the crucial variables that affect student participation and learning in asynchronous discussions are the presence of the instructor, deadlines, and the frequency of feedback. Schrire et al. studied the discussion/interaction from the multiple viewpoints of communication, cognition,

and discourse analysis to build a scheme for assessing knowledge building in asynchronous discussion groups [11]. SNA techniques combined with qualitative content analysis were used to study the impact of the teaching-learning process in higher education, with the resulting inference that, to be effective, online discussions must be closely related to learning goals and course objectives [12]. Dringus et al. studied many time-related participation indicators using data mining and text mining approaches that helped to extract temporal information from a threaded discussion. Nevertheless, sophisticated theoretical and empirical content analysis instruments are needed to improve the quality of $CSCL$ research by appropriately assessing well-defined, useful, and extractable discussion forums [13]. De Wever et al. presented an overview of different content analysis instruments and identified the need to develop a theoretical and empirical base of existing tools to help improve the overall quality of $CSCL$-research [14] [15].

Kim [16] analyzed the influence that the size of the discussion group has on the level of interaction of online student discussion forums using a case study methodology in an online course for the entire semester. This case study analysis revealed enhanced interactions in small group discussions compared to class-wide discussion forums. Thomas et al. studied the implications of using nonlinear branching structure of typical discussion forums in students' learning and revealed that efficient interface design is necessary to support more consistent many-to-many interactions [17]. They also suggest constructing a three-dimensional 'concept map' to facilitate unique opportunities for the students to understand the complex relationships in academic knowledge domains more deeply.

Since visualizing online discussions is critical for an active collaboration experience, tools have been developed to visualize the characteristics of the participants in a collaboration network, especially in students' discussion forums. These tools provide an efficient way to analyze students' interactions in asynchronous discussion forums using graph/network and Social Network Analysis (SNA) measures. For instance, a prototype system called ConverSpace $(ConversationSpace)$ utilizes an innovative interface to assist students during collaborative learning [9]. Sujana Jyothi et al. proposed a visualization tool for analyzing online discussions to review and support active online discussions of appropriate interventions every day. $Meerkat-ED$ is a toolbox that helps to analyze student activities using $CSCL$ tools. The tool can help visualize overall snapshots of participants in the discussion forums, their interactions, and the leaders versus peripheral students in the discussions. It also facilitates an abstract view of the hierarchical summarization of the topics discussed and measures the centrality of the students with respect to the number of posts and responses on specific discussion topics [18].

Graph data mining techniques have been widely used to analyze and visualize the patterns of interactions in collaboration networks. Girvan et al. investigated the structure of the community in networks of various kinds using centrality indices and information on edge betweenness to detect periphery of the community. Newman proposed an efficient algorithm to extract community structure and is much faster (typically thousands of times faster) than previous algorithms [19] [20]. SNA helps identify the significant people, teams and stakeholders who play a central role, determine information breakdown, bottlenecks, and structural holes, as well as isolated individuals, groups, and units [21] [22]. Hanneman and Riddle stated that SNA could be applied to a wide range of business problems, including knowledge management and collaboration, team building, human resources, sales and marketing, etc. [23]. Cross et al. [24], Ju Xiang, et al. [25], and Tang and Liu [26] studied SNA to understand the social (collective) behavior of a group (cluster or community) based on the relationship of the members in a group. While most of the real-time applications use relational data models for representing the transaction data, it is possible to gather additional knowledge by using sophisticated graph models where the inter-transaction dependencies are represented efficiently [27].

Although several studies use various techniques and tools to analyze the contents and ties (links) in asynchronous discussion forums, only very few studies focus on using structured discussion boards that contribute to better learning. The authors investigated collaborative learning strategies during a research study of a web-based learning tool that helps instructors integrate testing concepts into their CS courses. A sophisticated network-based analysis of student interaction patterns in course discussions exemplified the effective contribution of female students to STEM courses despite the low ratio of female students [28] [29]. Empirical research revealed that, beyond their direct effect on student learning in a particular course, collaborative learning techniques might offer broader benefits to *student-belonging* that would improve both course outcomes and retention in STEM majors [36] [37]. The need for many multi-faceted investigations was identified to create self-organizing and structured discussion boards to reinforce the voluntary participation of many students.

Many tools are developed for analyzing students' participation in collaborative learning using asynchronous discussion forums. These tools provide a graphical representation of student activities as a graph network and help teachers evaluate them. These tools also provide an idea of the patterns of interaction between students and their classmates. Some of the tools developed by the researchers across universities are discussed in this section.

Popolov et al. [30] developed a program ConverSpace(ConversationSpace), which is a prototype system that provides a spatial representation of computer-mediated conversation and provides assistance to students in the collaborative learning process. During a conversation, a user could highlight a text fragment of an existing remark or the whole remark and add a reply remark, which would automatically be linked to the first one. The newer version of this tool represents conversations on a 2D surface where the two axes denote time and topical structure. Feedback on this

tool showed that the spontaneity of discourse was restored and an increase in the speed of interaction was achieved. The tool could be used for both informal conversation or collaborative media.

A toolbox named Meerkat-ED was developed that created a hierarchical summarization of interactions in asynchronous discussion forums using Moodle for online courses. For this, the authors used SNA to extract the structure of students' interaction networks and community mining to identify the topics discussed by constructing a term network. Meerkat-ED calculated the degree centrality of the nodes (students) whose size represented their centrality and assisted in analyzing the leadership qualities of the students. Moreover, after a node was clicked, the messages from that student were displayed. The size of the edges represented the weight of communication between two nodes, and clicking on it displayed the messages between them. For the term network, the authors used the OpenNlp toolbox to extract noun phrases from discussions to represent nodes and used the FastModularity algorithm as the community detection algorithm [31].

Another tool, StudentViz, was developed to analyze student communication over social media tools(blog, wiki, microblogging tools) using eMUSE social learning environment to visualize the collaboration of communities and individual students. A data acquisition and graph building module, DtoG, was used to convert the data source into social graphs that could be visualized and SNA measures could be applied. The authors recorded posts and comments and calculated both graph theory metrics (in-degree, out-degree) and SNA centrality metrics (closeness, betweenness, eigenvector centrality, PageRank). Nodes were colored according to the created groups or communities and link direction was colored on the basis of the source node. The size indicated the rank of the node and the thickness of the links indicated the intensity of collaboration between two students [32]. The Virtual Interaction Mapping System(VIMS) tool helps visualize discussions using Moodle VLE within a webpage with hyperlinked nodes. [9].

The tools developed for visualizing and analyzing discussion forums generate vital information about the student network, whether it is showing the graph of students' interaction or extracting important details about the participants of the network. The development of these tools shows the potential of representing asynchronous or threaded discussion forums as a structured discussion. Although the above-mentioned tools focus on only one aspect of discussion or collaboration networks, nevertheless, support the argument that directly analyzing traditional discussion forums is difficult due to the limitation of its structure. Therefore, there is a need to extend the argument and develop a tool which could use all algorithms and measures discussed in this article.

B. Challenges in Modeling Student Discussion Boards and use of Graph Theoretic Concepts

The key challenges in network data analyses using the discussion board-based knowledge sharing scenario include the network modeling of data obtained from various data sources, determining metrics that describe the interaction patterns of the data elements or entities, and finding the community structure by using efficient algorithms. *Graphs* offer a robust framework based on information theory and graph theory to examine the statistical dependence between the entities under study and the relationships between these entities using appropriate similarity measures.

Typically, a discussion forum created by a facilitator for a course contains multiple discussion boards, each related to a separate topic. Each discussion board contains threads of discussion posts created by students related to a specific topic of a particular course assignment. First, we construct a collaborative network from the unstructured linear discussion board thread posts created by various students and investigate the working relationships of the students in the system using network measures. The graph/network model for each of the discussion thread posts is accomplished by considering the pairs of students S_i and S_j participating in a dialogue by posting their views as $nodes(vertices)$, and the association or connectivity between the two nodes as an $edge(link)$.

In weighted networks, the weights on the edges between two nodes are computed using the number of interactions between the two students or the type of information exchanged such as comments, views, suggestions, answering peer's questions, shared challenges, solutions to a question, etc. Thus, the resulting graphs can be modeled as weighted and undirected (or directed) networks representing the student discussion boards. These networks can be used to compute various network measures that are used to characterize the networks. Inferential statistical methods are used for further analysis.

C. Network Analysis Measures - Definitions

The following subsections describe the important terms and measures used in this paper to characterize the student discussion network.

1) Connected Component: A *connected component* of an undirected graph G is a subgraph in which every node in the subgraph has a path consisting of a sequence of edges from that node to any other node in the subgraph and for which there is no path to any graph G not in this subgraph.

2) k-Core: A $k-core$ of a graph G represents a connected component of G where all nodes have a degree of at least k.

3) Degree Centrality: Degree Centrality (DC) determines the immediate influence of a node in terms of the number of direct connections to it [33]. We can represent a graph as an adjacency matrix *A* where A_{ij}, $1 \leq i, j \leq n$, is the element in the i^{th} row and j^{th} column with the value 1 or 0 based on whether there is or is not a link between the nodes *i* and *j*. (Note that in an undirected graph, this is a symmetric matrix.) The DC of a node i is the sum of direct interactions with nodes j that can be calculated as shown in (1).

$$DC(i) = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{A_{ij}} \tag{1}$$

Nodes with high degree centrality are considered prominent nodes $(hubs)$ that serve as a source of information exchange with many other nodes. The Average Degree Centrality

(ADC) of a network is the average of DCs of all nodes in the network. Higher ADC indicates better interconnectivity among the nodes in the network.

4) Clustering Coefficient: The *clustering coefficient* CC is a measure of the degree to which the nodes of a graph tend to cluster together. We let e_i be the *degree* of node i (that is, e_i is the number of nodes adjacent to node i or the number of edges connected to that node i.) Then the CC of a node i in a network with n nodes is the ratio of e_i to the maximum possible number of edges between those neighbors, $\frac{n(n-1)}{2}$ and is calculated as shown in (2).

$$CC(i) = \frac{2e_i}{n(n-1)} \tag{2}$$

In this paper, we focus on creating collaboration networks from student discussion board data and applying state-of-the-art techniques for mining and identifying student characteristics (traits) and evaluating the impact that collaborations have on student course performance. Our study of patterns of participation and discourse in online discussions resulted in an important finding that, although the density of interactions between participants was high, each discussion thread itself was short. The discussion boards were a linear collection of posts rather than a graph-based representation making it difficult for students to visualize and respond to the variety of discussions posted by peers. We address the need for structured, network schema-based, and moderated discussion forums to facilitate more extensive, and thus more effective exchanges between the participants rather than short linear threads of discussions.

III. Methodology

Online discussion boards, generally organized as tree structures known as threads, remain a rich repository of student-generated content. The purpose of this study is to gather insights into students' sense of *belongingness* in computer programming courses. First, we examine how the corpus of student discussion threads in LMS can be efficiently modeled as a collaboration network using graph theoretical concepts. We then demonstrate the impact that this collaboration network has created on students' performance in STEM courses.

A. Research Hypothesis

Our working research hypothesis is that structured discussion forums facilitate the drawing of inferences from the collaboration network in ways that are not possible with unstructured forums. A framework was developed, experiments were conducted, and some empirical results are presented to prove this hypothesis. To address the limitations of drawing inferences in the existing threaded discussions, we identified the need for a tool that could aid exploratory analysis in a student collaboration network. Therefore, we address the research hypothesis by representing discussion boards using a schema-based structured network model. This could help students participate interactively in discussions and share knowledge. On the other hand, it would provide teachers with a platform to analyze students' performance using the results generated from this tool. The proposed framework based on the research hypothesis is shown in Figure 1.

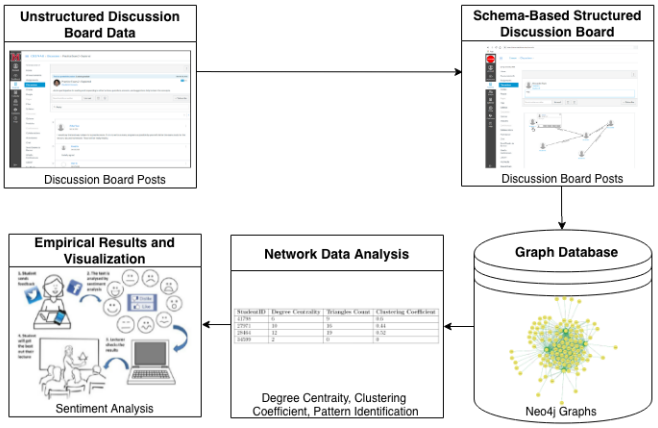

Fig. 1. Proposed Framework for Schema-Based Structured Discussion Board

It shows the proposed design of the structured schema-based discussion forum tool. Once a student writes a post in the discussion board as a response to another student's post, a node and an edge connecting the source and the destination nodes appear as shown. It helps in both encouraging the students' participation and effective visualization of the discussion board, in addition to using a schema-based approach to store the discussion data.

The framework we used in our link analysis of the discussion board data collected from the collaboration network is shown in Figure 2.

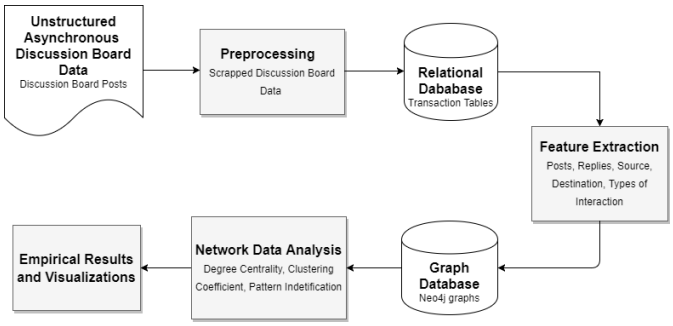

Fig. 2. Framework for Link Analysis in Goal-Oriented Collaboration Networks

B. Data Collection

Data used in this study were collected from $Fall$ 2016 and $Fall$ 2017 semesters of an Introduction to Programming course $(CS1)$ at a medium-sized public university. Both of these courses were handled by the same faculty and the same method of instruction was followed. $Fall$ 2016 data collected from 102 students was used as a control group; Students in that semester were not involved in a formal knowledge sharing activity based on the discussion board. We used the final grades of these students as benchmark data. $Fall$ 2017 Semester data from 129 students represent the treatment group

where discussion boards served as a platform for students to share their learning experiences during the course by posting their opinions on different discussion topics and questions related to the information presented in the course [29]. The analysis included data from the online discussion forum and the final grades of the treatment group; for the control group, only the final grades were considered for the analysis.

The discussion forum contains seven online discussion boards, DB_1 to DB_7, each implemented on the $Canvas\ LMS$ used to support the course. A discussion board was created after the mid-term exam to capture student questions and concerns; other discussion boards were tied to programming assignments. These were open for seven days after the assignment date to allow students to post their ideas, new things they have learned that are applicable or challenges faced. The student participation was worth 10% of the overall course grade (total participation points = 397 points + 14 points for Discussion Boards). Thus, the discussion posts written by the students were graded as 3.5% of the overall grade of the course. Points were awarded based on the type of discussion posts. For example, sharing challenges is given 2 points, and comments on others' posts are given 0.5 points. The data collected from the discussion posts were modeled as graphs after preprocessing the unstructured data from the discussion posts. Please refer [37] for details about the preprocessing of data.

Each transaction includes the details of the source entity S_j (student that initiates the discussion post), destination entity S_k or DB_l, the discussion board name, and the type of discussion. The destination entity may be another student S_k or the discussion board itself DB_l, The transaction table created for this study contains 1044 distinct discussion records. Table I presents an example of how data are stored in the transaction table. Notice that the student S_4 posted a solution to the student S_2 on the discussion board DB_2 while the student S_2 posts a comment on the discussion board DB_2 itself. The table does not show all information used for the analysis, such as dates of the posts, discussion tokens (keywords), etc.

TABLE I
DISCUSSION BOARD TRANSACTION TABLE

TID	Source	Destination	DB Name	Discussion Type
1	S1	S4	DB_1	comments
2	S1	S3	DB_1	discusses
3	S2	DB_2	DB_2	comments
4	S4	S2	DB_2	solves
5	S4	S1	DB_3	share challenge

Each transaction (discussion) post is classified into one of the seven types of information that it contains as discussed above. We also compute the number of posts made by each student under each type of discussion in the discussion forum. In addition to all these data collected from the treatment group, the final grades of students in both treatment and control groups were used for the study. The final grades of the students in the control group serve as baseline data to compare overall performance. It should be noted that students from the control group were not explicitly assigned any online discussion forums as part of the instruction. The results of the experiments and the detailed discussion are presented in the following section.

IV. RESULTS AND DISCUSSION

This section presents the work plan and the experiments performed with a focus on answering the research question.

A. Discussion Board User Interface

We first model the threads of unstructured asynchronous discussion board post data as a weighted undirected graph. The discussion board user interface tool design includes a graphical representation of Canvas LMS discussion board based on a schema, which represents the discussion board as a graph where the participants (students and teachers) and the discussion board are vertices (nodes), and the posts and types of replies between the participants are the edges (relationships) connecting those nodes. The website is developed using Web 2.0 technology and is embedded in the canvas course menu as an additional app. Clicking on this app redirects users to the website url. The development and use of this tool are explained in the following sections.

B. Overview of Schema-based Model

The Discussion Board User Interface tool appears as an external app in Canvas that opens in a new window. This tool consists of two components, Front End and Back End. The discussion board data on Canvas is fetched using the back end and visualized in the application window as a directed graph with the help of the front end. The discussion board user interface architecture is given in Figure 3.

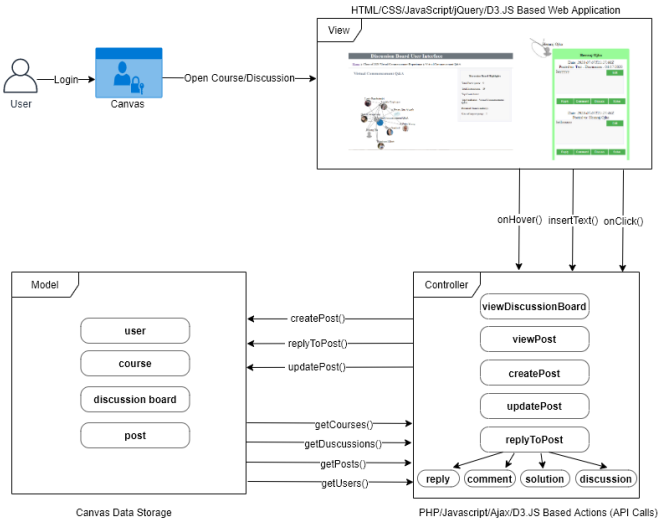

Fig. 3. Discussion Board User Interface Architecture

C. Discussion Board Front-End

A list of active courses appears and selecting one of them shows the discussion boards available for that particular course. Selecting a discussion displays the graph where

students and the discussion topic appear as nodes and the posts appear as links between the nodes. Multiple edges between two nodes represent multiple posts shared between two nodes. New nodes and links are created when a new user adds a post. A post can be one of four categories, namely,'reply','question','solve,' and 'discuss'. An existing post by a user can be updated by that user, but cannot be deleted. The teacher (instructor) has access to update every post in the discussion. Upon visualization of a discussion, discussion highlights appear, which use Canvas LMS data, SNA metrics, and graph algorithms to show important information about a discussion. The features of the Discussion Board User Interface Tool include the downloadable CSV file that helps teachers to grade student participation, a wordcloud of important and frequent words in the discussions, and the gamification element of the leaderboard that displays different titles of participants that achieve any specific milestones like top contributor, top facilitator, top friend and potential team leader are discussed in the following subsections.

D. Discussion Board Back-End

This subsection provides detailed information about the programming languages, APIs, libraries, and packages used for the development of this tool.

1) Data Collection and Storage: The data captured by this tool are fetched from Canvas LMS using REST API [34]. A REST or RESTful API is an application program interface (API) which makes use of HTTP requests to GET, PUT, POST, and DELETE data. All data is stored in the Canvas LMS server. GET is used to retrieve a resource. PUT is used to change the state of or update a resource. POST is used to create a resource. And DELETE is used to remove a resource. An API for a website is a code that helps two software programs to share data and information with each other [35]. This discussion board user interface uses mainly PHP and Ajax scripts to perform operations on Canvas LMS REST API. The endpoints for REST operations are: List Courses, Discussion Topics, and Entries/Replies; Post a Discussion Entry and Reply; and Update Discussion Entry/Reply.

2) Data Visualization and Manipulation: The data collected from the canvas database is received as JSON data and then configured on the website using HTML, CSS, JavaScript, jQuery, and AJAX.

E. Discussion Board User Interface Tool

Discussion Board User Interface Tool is integrated into Canvas with the help of Redirect Tool that is developed by Instructure. Upon integration, the Discussion Board User Interface can be seen in the course navigation section on the left of the Canvas LMS window.

F. Tooltips

Tooltips are present on the nodes, which, when clicked, display a list of posts made by that participant. The user can edit his/her posts by clicking on 'edit' button. For a teacher, he/she can edit everyone's posts in that discussion. Users can reply to any node by clicking on one of the four buttons, reply, question, discuss, or solve present below each post. A snapshot of the tooltip is shown in Figure 4.

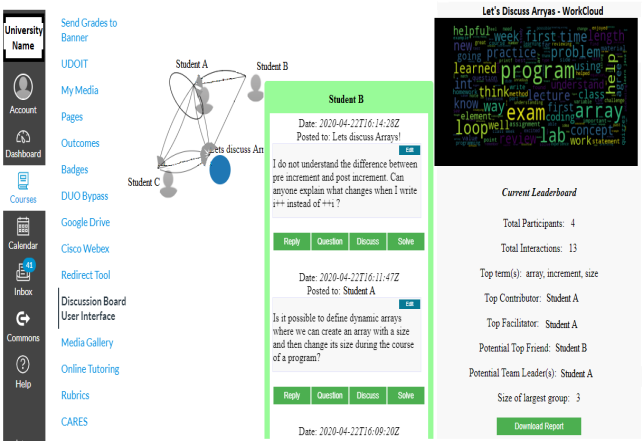

Fig. 4. Discussion Board User Interface in Canvas and Node Tooltip

G. Discussion Board Highlights and Wordcloud

When a discussion graph is displayed, it also shows important features of the network on the right side of the application window. The wordcloud of all posts in the current discussion is displayed. The words that occur more frequently are larger in size than the words that occur less frequently (see Figure 4).

H. Current Leaderboard

Current leaderboard adds a gamification feature to this tool. It contains values as follows:

- **Total Participants:** The total number of participants in the discussion is calculated by counting the number of nodes in the graph that exclude the discussion node.
- **Total Interactions:** The total number of interactions in the discussion is calculated by counting the number of links in the graph.
- **Top Term(s):** The top three terms that occur most frequently are taken from the Wordcloud database, if there are any.
- **Top Contributor:** The participant node (except discussion topic node) which has the highest number of outgoing links (out-degree centrality) is termed as the top contributor.
- **Top Facilitator:** The participant node(except discussion topic node) which has the highest betweenness centrality is termed as the top facilitator as that participant is a node in between that facilitates other nodes to contribute.
- **Potential Friend:** The participant node (except discussion topic node) which has highest clustering coefficient is termed as the top contributor as that participant has the tendency to cluster and create groups.
- **Potential Team Leader:** The participant node(except discussion topic node) which has the highest average centrality is termed as the top contributor. The out-degree centrality, betweenness centrality, and clustering

coefficients are averaged out, and the participant's name with highest score is displayed.
- **Largest Group Size:** Size of the largest group in the discussion is the size of the largest clique. It represents a close group where all nodes interact with each other more than any other group in the network.

I. Downloadable CSV Report

The CSV (Comma Separated Values) file containing participants' id, participants' name, normalized degree centrality (DC), betweenness centrality(BC), clustering coefficient(CC), triangle count (TC) and cluster label for clusters they belong to (if any) (CL),etc. can be downloaded as shown in Table II.

TABLE II
A SAMPLE CSV REPORT

ID	Name	Posts	DC	BC	CC	TC	CL
80760	Student A	7	11	0.167	0.667	2	3
81680	Student B	3	5	0	1	1	0
90722	Student C	3	4	0	1	1	1
874525	Discussion 1	1	6	0.167	0.667	2	2

J. Student's Discussion Statistics/Details

This subsection is only visible to teachers who need to grade the students who participate in the discussion. The results are seen on the right hand side of the window discussion board highlighted once a node is clicked. The following results are included in the statistics:

- **Interactions Frequency** - total number of posts made by the participant.
- **Interaction List** - List of other participants in the discussion with which the selected participant has interacted.
- **Type of Replies** - Count of replies of all four types (reply, question, discuss, solve) of a participant in the discussion.

The new structured, schema-based discussion board tool provided a visualization of students' interactions (a directed graph), where one could easily see the interactions between people in the network. The lists of posts made by a participant are presented with a click on the node. The use of various algorithms for analyzing the features of the discussion board has made it more result oriented. The teachers can see how a student has performed in a discussion and can also fetch the required statistical results for grading a student as well as understanding the level of their knowledge acquisition.

V. THREATS TO VALIDITY

The goal of this research project is to gather insights into students' sense of belongingness in computer programming courses, which of course is a sense of belongingness and is a complex emotional characteristic that is difficult to assess directly. However, we believe that our measurement of student interaction via discussion boards is a reasonable first-order approximation for assessing that.

A potential threat to the validity of this study is the limited study size. A control group of 102 students participated in the course one semester, and a treatment group of 129 students participated in the same course one year later. From a statistical perspective, the number of participants and database records produced is reasonable. It is, however, necessary to validate the result in the context of other universities and professors. Furthermore, this study was conducted within the context of Canvas as a learning management system. Porting this application to another learning management system would not be difficult since they both have similar functionality.

VI. CONCLUSION AND FUTURE WORK

This research work analyzes the interaction patterns of the discussion boards available in Canvas LMS, which can help instructors analyze and measure the participation of students and their belongingness to the course. To address the problem of asynchronous and threaded structure of discussion boards, we came up with a new schema-based discussion board model that graphically models the discussion boards and makes discussions more interactive and result-driven. The new discussion board tool is schema-based and structured and includes gamification elements such as the leaderboard that allow students to share information and see their participation as measured using various SNA metrics. The teachers on the other hand, can easily grade the discussion in real time and see interesting results about the interactions which, otherwise, are not available in threaded discussion forums. Therefore, this new research model has the potential to support student-centered learning engagement and improve belonging in classrooms.

Since this schema-based structured discussion forum is a working prototype, this work can be extended by adding new functionalities that help gather more insights about the discussions. Students could be grouped into communities and modeled on similar graphs, which would help to observe the behavior of students. Feature extraction would provide additional information about the discussions such as displaying the important questions discussed and listing their possible solutions,etc. Moreover, all discussions could be combined as a single graph for the whole course. The individual discussion graphs for a course displayed in specific discussions can be aggregated together for all courses the students are enrolled to gather full insights (e.g., temporal analysis) about the student's interaction patterns and overall knowledge acquisition. To make it more interesting for students, additional gamification elements like rewards and points table can be added.

As per the efficiency of the discussion board, the user interface tool is concerned, although the student or teacher is automatically authenticated into the tool once he/she logs into Canvas LMS, the loading time can be reduced and the script takes some time to load. The application can be made more secure by adding *OAuth* authentication. The data used in this application is not stored anywhere else but in canvas, which affects the speed of this tool in case of any extra calculations like type of reply to a discussion entry is performed. The application takes a while to reload the graph when an update is made, which can also be avoided by making the updates more dynamic.

REFERENCES

[1] Prokess, Ann M., and Anna McDaniel. "Are nursing students engaged in learning? A secondary analysis of data from the National Survey of Student Engagement." Nursing education perspectives 32, no. 2 (2011): 89-94.

[2] Salter, Nicholas P., and Marissa R. Conneely. "Structured and unstructured discussion forums as tools for student engagement." Computers in Human Behavior 46 (2015): 18-25.

[3] Reffay, Christophe, and Thierry Chanier. "Social network analysis used for modelling collaboration in distance learning groups." In International Conference on Intelligent Tutoring Systems, pp. 31-40. Springer, Berlin, Heidelberg, 2002.

[4] Hiltz, Starr Roxanne. "Collaborative Learning in Asynchronous Learning Networks: Building Learning Communities." (1998).

[5] Ramasamy, Vijayalakshmi, James D. Kiper, Hemraj Ojha, and Urvashi Desai. Analyzing Link Dynamics in Student Collaboration Networks using Canvas-A Student-Centered Learning Perspective. In 2019 IEEE Frontiers in Education Conference (FIE), pp. 1-9. IEEE, 2019.

[6] Fluegeman, R.H., and S. Rice, Maintaining a small-group discussion focus while bringing international issues into the large classroom. Journal of Geoscience Education, Vol. 52 (3), pp. 260–65, 2004.

[7] Wilen, W, Refuting misconceptions about classroom discussion, The Social Studies, Vol. 95 (1), pp. 33–40, 2004.

[8] Johnson, J.P., and A. Mighten, Research Briefs - A Comparison of Teaching Strategies: Lecture notes combined with structured group discussion versus lecture only, Journal of Nursing Education, Vol. 44 (7), pp. 319–22, 2005.

[9] Sujana Jyothi, Claire McAvinia, John Keating, A visualization tool to aid exploration of students' interactions in asynchronous online communication, Computers & Education, vol. 58, pp. 30-42, 2012

[10] Dennen, V. P., From Message Posting to Learning Dialogues: Factors affecting learner participation in asynchronous discussion, Distance Education, Vol. 26, pp. 127-148, 2005.

[11] Sarah Schrire, Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis, Computers & Education, Vol. 46, pp. 49-70, 2006.

[12] Erping Zhu, Interaction and cognitive engagement: An analysis of four asynchronous online discussions, Instructional Science, Vol. 34, pp. 451-480, 2006.

[13] Laurie P. Dringus, Timothy Ellis, Using data mining as a strategy for assessing asynchronous discussion forums, Computers & Education, Vol. 45, pp. 141-160, 2005.

[14] De Wever, B, Schellens, T, Valcke, M, Van Keer, H, Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review, Computers & Education, vol. 46, pp. 6-28, 2006.

[15] N. Du, B. Wu, X. Pei, B. Wang and L. Xu, "Community Detection in Large-Scale Social Networks," Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, August 2007, pp. 16-25.

[16] Jungjoo Kim, Influence of group size on students' participation in online discussion forums, Computers & Education, Vol. 62, pp. 123-129, 2003.

[17] Thomas, M.J.W, Learning within incoherent structures: the space of online discussion forums, Journal of Computer Assisted Learning, vol. 18, pp. 351-366, 2002

[18] Reihaneh Rabbany, Mansoureh Takaffoli, Osmar R. Zaiane, Social Network Analysis and Mining to Support the Assessment of Online Student Participation, ACM SIGKDD Explorations on Educational Data mining, December 2011

[19] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences, vol. 99, pp. 7821-7826, 2002

[20] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E, vol. 69, Issue 2, 2004

[21] Dorogovtsev, S.N. and Mendes, J.F.F. "Evolution of networks", Advances in Physics, Vol. 51, pp. 1079-1187, 2002.

[22] [Anonymous 2011]

[23] Hanneman, R.A. and Riddle, M. "Introduction to social network methods", CA: University of California, Riverside, CA: University of California, Riverside, http://faculty.ucr.edu/ hanneman, 2005.

[24] Cross, R., Borgatti, S.P. and Parker, A. "Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration", California Management Review, Vol. 44, pp. 25-46, 2002.

[25] Xiang, J., Hu, K. and Tang, Y. "A class of improved algorithms for detecting communities in complex networks ", Physica A: Statistical Mechanics and its Applications, Vol. 387, pp. 3327-3334, 2008.

[26] Tang, L. and Liu, H. "Community Detection and Mining in Social Media", Morgan & Claypool Publishers, 2010.

[27] Ivan F.Videla-Cavieres Sebastián A.Ríos. Extending market basket analysis with graph mining techniques: A real case, Vol. 41, pp. 1928-1936, 2014.

[28] Vijayalakshmi Ramasamy, Hakam Alomari, James Kiper and Geoffrey Potvin. A Minimally Disruptive Approach of Integrating Testing into Computer Programming Courses, Second IEEE/ACM International Workshop, Gothenburg, Sweden, June 2, 2018.

[29] Vijayalakshmi Ramasamy, U. Desai, H. W. Alomari and J. D. Kiper, TP-GraphMiner: A Clustering Framework for Task-Based Information Networks, IEEE International Conference on System, Computation, Automation and Networking (ICSCA), pp. 1-7, 2018.

[30] Popolov, Dimitri, Callaghan, Michael, and Luker, Paul, Conversation Space: Visualising Multi-threaded Conversation, Proceedings of the Workshop on Advanced Visual Interfaces, pp. 246-249, 2000.

[31] Rabbany, Reihaneh, Mansoureh Takaffoli, and Osmar R. Zaïane. "Analyzing participation of students in online courses using social network analysis techniques." In Proceedings of educational data mining. 2011.

[32] Becheru, Alex, Andreea Calota, and Elvira Popescu. "Analyzing students' collaboration patterns in a social learning environment using studentviz platform." Smart Learning Environments 5, no. 1 (2018): 1-18.

[33] L.C. Freeman, Centrality in Social networks: Conceptual Clarification, Social Networks, 1, 215-239, 1979.

[34] Instructure, Inc.,"Canvas LMS REST API Documentation", 2020, https://canvas.instructure.com/doc/api/.

[35] Gillis, Alexander S. "What is REST API (RESTful API)?" SearchAppArchitecture, Techtarget, 22 September 2020, https://searchapparchitecture.techtarget.com/definition/RESTful-API.

[36] Urvashi Desai, Vijayalaskhmi Ramasamy, Kiper James. "A Study on Student Performance Evaluation Using Discussion Board Networks." Proceedings of the 51st ACM Technical Symposium on Computer Science Education, pp. 500–506, 2020.

[37] Vijayalakshmi Ramasamy, Kiper James, Ojha Hemraj, and Urvashi Desai. "Analyzing Link Dynamics in Student Collaboration Networks using Canvas-A Student-Centered Learning Perspective", IEEE Frontiers in Education Conference (FIE 2019), 2019.