

Social Network Analysis and Mining to Support the Assessment of On-line Student Participation

Reihaneh Rabbany k., Mansoureh Takaffoli and Osmar R. Zaiane

Department of Computing Science

University of Alberta

Edmonton, Canada

rabbanyk,takaffol,zaiane@ualberta.ca

ABSTRACT

There is a growing number of courses delivered using e-learning environments and their online discussions play an important role in collaborative learning of students. Even in courses with a few number of students, there could be thousands of messages generated in a few months within these forums. Manually evaluating the participation of students in such case is a significant challenge, considering the fact that current e-learning environments do not provide much information regarding the structure of interactions between students. There is a recent line of research on applying social network analysis (SNA) techniques to study these interactions.

Here we propose to exploit SNA techniques, including community mining, in order to discover relevant structures in social networks we generate from student communications but also information networks we produce from the content of the exchanged messages. With visualization of these discovered relevant structures and the automated identification of central and peripheral participants, an instructor is provided with better means to assess participation in the online discussions. We implemented these new ideas in a toolbox, named Meerkat-ED. Which automatically discovers relevant network structures, visualizes overall snapshots of interactions between the participants in the discussion forums, and outlines the leader/peripheral students. Moreover, it creates a hierarchical summarization of the discussed topics, which gives the instructor a quick view of what is under discussion. We believe exploiting the mining abilities of this toolbox would facilitate fair evaluation of students' participation in online courses.

1. INTRODUCTION

There is a growing number of courses delivered using e-learning environments, especially in postsecondary education, using computer-supported collaborative learning (CSCL) tools, such as Moodle [20], WebCT [30] and Blackboard [1]. Online asynchronous discussions in these environments play an important role in collaborative learning of students. It makes them actively engaged in sharing information and perspectives by interacting with other students [12].

There is a theoretical emphasis in CSCL on the role of threaded discussion forums for online learning activities [2]. Even basic CSCL tools enable the development of these

threads where the learners could access text, revise it or reinterpret it; which allow them to connect, build, and refine ideas, along with stimulating deeper reflection [2]. There could be thousands of messages generated in a few months within these forums, containing long discussion threads bearing many interactions between students. Therefore the CSCL tools should provide a means to help instructors for evaluating participation of students and analyzing the structure of these interactions; which otherwise could be very time consuming, if not impossible, for the instructors to be done manually.

Current CSCL tools do not provide much information regarding the participation of students and structure of interactions between them in discussion threads. In many cases, only some statistical information is provided such as frequency of postings, which is not a sufficient or even useful measure for interaction activity [12]. This means that the instructors who are using these tools, do not have access to convenient indicators that would allow them to evaluate the participation and interaction in their classes [31]. Instructors usually have to monitor the discussion threads manually which is hard, time consuming, and prone to human error.

On the other hand, there exists a large body of research on studying the participation of students in such discussion threads using traditional research methods: content analysis, interviews, survey observations and questionnaires [11]. These methods try to detect the activities that students are involved in while ignoring the relations between students. For example, content analysis methods, as the most common traditional methods, provide deep information about specific participants. However, they neglect the relationships between the participants while their focus is on the content, not on the structure [31].

In order to fully appreciate the participation of students, we need to understand their patterns of interactions and answer questions like who is involved in each discussion, who is the active/peripheral participant in a discussion thread [11]. Nurmela et al. [24], Haythornthwaite [15] and Cho et al. [6] demonstrated the practicality of social network analysis methods in CSCL, as a method for obtaining information about relations, fundamental structural and collaborative patterns. Moreover, there is a recent line of work on applying social network analysis techniques for evaluating the participation of students in online courses e.g. Dawson et al. [9], Sundararajan [28], Calvani et al. [2], de Laat et al. [11], Willging [31], Laghos and Zaphiris [17], Erlin et al. [12]. The major challenges these attempts tried to tackle are: extracting social networks from asynchronous discussion forums

(might require content analysis), finding appropriate indicators for evaluating participation (from education's point of view) and measuring these indicators using social network analysis. As clarified in the related works, Section 2, none of these works provides a complete or specific toolbox for analyzing discussion threads. However, they attempted to address one of these challenges to some extent.

Here, we elaborate on the importance of social network analysis for mining structural data in the field of computer science and its applicability to the domain of education. We propose Meerkat-ED¹, a specific and practical toolbox for analyzing interactions of students in asynchronous discussion forums of online courses. Meerkat-ED analyzes the structure of these interactions using social network analysis techniques including community mining. It prepares and visualizes overall snapshots of participants in the discussion forums, their interactions, and the leader/peripheral students in these discussions. Moreover, it analyzes the content of the exchanged messages in this discussions by building an information network of terms and using community mining techniques to identify the topics discussed based on the word sense communities [4]. Meerkat-ED creates a hierarchical summarization of these discussed topics in the forums. This gives the instructor a quick view of what is discussed in these forums. It further illustrates how much each student has participated in these topics, by showing his/her centrality in the discussions on that topic, the number of posts, replies, and the portion of terms used by that student in the discussions.

In the following, we first introduce some basic concepts in social network analysis and elaborate on possible applications in the context of on-line Education. We then present Meerkat-ED – our solution for social network analysis of on-line courses in Section 3 and illustrate its practicability on our own case study data in Section 4.

2. BACKGROUND AND RELATED WORKS

Social networks are formally defined as a set of actors or network members which are tied by one or more types of relations. The actors are most commonly persons or organizations, however, they could be any entities such as web pages, countries, proteins, documents, etc. There could also be many different types of relationships, to name a few, collaborations, friendships, web links, citations, information flow, etc. These relations are represented by the edges in the network connecting the actors and may have a direction indicating the flow from one actor to the other; and a strength denoting how much, how often, or how important the relationship is.

2.1 Social Network Analysis

Unlike proponents of attribute based social sciences, social network analysts argue that causation is not located in the individuals, but in the social structure [19]. Social network analysis is the study of this structure.

Rooted in sociology, nowadays, social network analysis has become an interdisciplinary area of study, including researchers from anthropology, communications, computer science, education, economics, criminology, management science, medicine, political science, and other disciplines [19].

¹Meerkat-ED could be downloaded from:
<http://meerkat.aicml.ca/meerkat-ED>

Social network analysis examines the structure and composition of ties in a given network and provides insights into its structural characteristics, most commonly:

1. Central actors in the network (prestige)
2. Individuals with
 - (a) The most outgoing connections (influence)
 - (b) The most incoming connections (prominence)
 - (c) The least connections (outlier)
3. How connected are neighbours of an actor (cliquishness)
4. Actors that are communicating more often with each other (community)
5. Actors that are involved in passing information through the network (path length)
6. Proportion of possible ties that actually exist in the network (density)

Here we elaborate more on two of these structural characteristics which would be used latter in our analysis.

Centrality

Centrality of a node in a community measures its relative importance within that community. There are many measures of centrality, namely degree, betweenness, closeness and eigenvector centrality measures. For example the degree centrality for a node within a community represents to some extent the *popularity* of that node in the community and simply measures the number of edges from the community incident upon that node. While betweenness centrality represents the *control* of a node over communication within its community and measures the number of shortest paths passing through that node.

Community Mining

Detecting communities – densely connected actors – has been pursued by sociologists for many decades. More recently, it has also attracted attention from physicists, applied mathematicians and computer scientists [22]. This interest resulted in the emergence of a variety of different community detection approaches, e.g. Clique percolation by Palla et al. [26], FastModularity by Clauset and Newman et al. [7] and Local by Chen et al. [5] (refer to [18] for a recent survey).

Modularity optimization based approaches are the most prominent family of community detection methods. The modularity Q is proposed as a measure of the quality of a particular division of a network. The basic idea is to compare the division to a randomized network with exactly the same vertices and same degrees, in which edges are placed randomly without regard to community structure [23].

2.2 Social Network Analysis of Asynchronous Discussions in Online Courses

In order to apply social network analysis techniques to assess participation of students in an e-learning environment, we need to first extract the social network from the e-learning forums, then we must consider which measures show an effective participation, and finally report these measures in an appropriate way. Here, we give an overview of the previous works related to each of these phases.

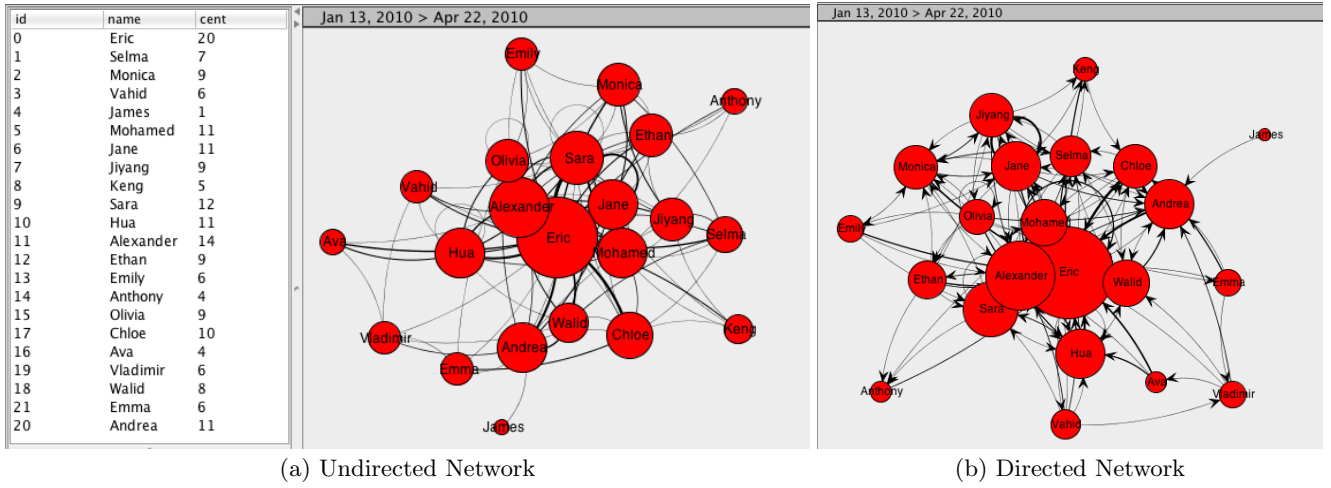


Figure 1: Visualized Student Network: The left panel lists the students in the course. The right panel shows the social network of interaction of students in the course. The size of nodes corresponds to their centrality/leadership in the discussions. The width of edges represents the weight of communication between incident nodes.

Extraction of Social Network

CSCL tools record log files that contain the detailed actions made by learners. Hence, log files include information about the activity of the participants in the discussion forums [24]. de Laat et al. [11], Willging [31], Erlin et al. [12] and Laghos and Zaphiris [17] used these log files to extract the social network underneath of discussion threads. Laghos et al. stated that they considered each message as directed to all participants in that discussion thread while others considered it as only directed to the previous message.

Gruzd and Haythornthwaite [14] and [13], proposed an alternative and more complicated way of extracting social networks, called named network. They argue that using this common method (connecting a poster to the previous poster in the thread) would result in losing much of the connections. Their approach briefly is: first using named entity recognition to find the nodes of the network, then counting the number of times that each name is mentioned in posts by others to obtain the ties, and finally weighting these ties by the amount of information exchanged in the posts. However, their final reported results are not that promising and even obtaining those results requires many manual corrections during the process.

More recently, Dawson et al. [9; 10] developed a cross platform toolbox called SNAPP which is able to capture the discussion threads of different CSCL platforms from their content in the web browser. However this crawling process is very time consuming comparing to reading an input log file.

Regarding what we should consider as the participation in extracting the social network, Hrastinski [16] suggested that apart from writing, there are other indicators of participation like accessing the e-learning environment, reading posts or the quantity and quality of the writing. However, all of these methods extracted networks just based on posts by student (writing level).

Measuring the Effectiveness of Participation

Daradoumis et al. [8] defined high level weighted (showing the importance) indicators to represent collaboration learning process; task performance, group functioning, social support, and help services. They further divided these indicators to skills and sub-skills, and assigned every sub-skill to an action. For example, group functioning is divided into: active participation behaviour, task processing, communication processing, etc. On the other hand, communication processing is itself divided into more sub-skills: clarification, evaluation, illustration, etc. and clarification is then mapped to the action of changing description of a document or url.

In the education context, Calvani et al. [2] defined 9 indicators for measuring the effectiveness of participation to compare different groups within a class; extent of participation (number of messages), proposing attitude (number of messages with proposal label), equal participation (variance of messages for users), extent of role (portion of roles used), rhythm (variance of daily messages per day), reciprocal reading (portion of messages that have been read), depth (average response depth), reactivity to proposal (number of direct answers to messages with proposal label) and conclusiveness (number of messages with conclusion label); all summarized in a nonagon graph which shows the group interactions relatively to the mean behaviour of all groups.

However, for measuring the effectiveness of participation, most of the previous works simply used general social network measures (different centrality measures, betweenness, etc.), available in one of the common general social network analysis toolboxes. Sundararajan [28], de Laat et al. [11], Willging [31], Erlin et al. [12] used UCINET [29], Laghos and Zaphiris [17] used NetMiner [21] and Dawson et al. [9; 10] developed their own SNA toolbox which offers simple visualization and limited analysis.

3. SOCIAL NETWORK ANALYSIS FOR EDUCATION: MEERKAT-ED

In this section, we illustrate the practicability of social network analysis in evaluating participation of students in online discussion threads. We present our specific social network analysis toolbox, named Meerkat-ED, to analyze online courses. Meerkat-ED is an educational purpose edition of Meerkat [3] social network analysis toolbox. Meerkat-ED is designed for assessing the participation of students in asynchronous discussion forums of online courses. It analyzes the structure of interactions between students in these discussions using social network analysis techniques. It exploits community mining techniques in order to discover relevant structures in social networks generated from student communications and also information networks produced from the content of the exchanged messages. With visualization of these discovered relevant structures and the automated identification of central and peripheral participants, an instructor is provided with better means to assess participation in the online discussions.

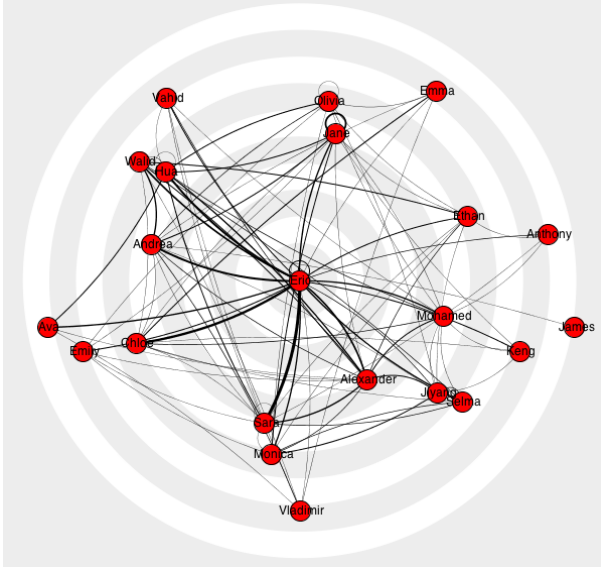


Figure 3: Comparing centrality of students: the students closer to the center are more central in the student network, i.e., have participated more in the discussions of the course. Likewise, the further from the center, the less the student was active; here James is the least active student in the discussions and is placed on the outer circle.

Meerkat-ED prepares and visualizes dynamic snapshots of participants in the discussion forums, their interactions, and the leader/peripheral students. For any specific period of time, it creates a hierarchical summarization of the topics discussed in the forums using community mining. Which gives the instructor a quick view of what is under discussion in these forums. It further illustrates how much each student has participated on these topics, by showing his/her centrality in the discussions on that topic, the number of posts, replies, and the portion of terms used by that student in discussions on the topic. Meerkat-ED builds and analyzes two kinds of networks out of the discussion forums: a *social network of the students* where links represent cor-

respondence, and a *network of the phrases* used in the discussions where links represent co-occurrence of phrases in the same sentence. Interpreting the first network shows the interaction structure of the students participating in the discussions. Furthermore, centrality of students in this network corresponds to their leadership in the discussions. Interpreting the network of phrases depicts the phrases used in the discussion and the relations between these phrases. Finding the hierarchical communities in this network demonstrates the topics addressed in the discussions. Choosing each of these topics outlines the students who participated in that topic and the extent of their participation.

3.1 Interpreting Students Interaction Network

Interpreting the network of interaction between students helps instructors monitor the interaction structure of students, and examine which students are the leaders in given discussions and who are the peripheral students. Here, we first describe how the network is extracted based on the information from the discussion threads. Then, we continue by bringing an analysis of leadership of the students based on their centrality in this network.

Overall Structure of the Interactions

The student network shows the interaction between students in the discussion forums, where the nodes represent students of the course and edges are the interaction between these students (i.e. messages exchanged). The edges are weighted by the number of messages passed between the two incident students. This network could be built both directed or undirected (chosen by the instructor); in the directed model, each message is considered connecting the author of the message to the author of its parent message.

Detecting Influential Students

The leadership and influence of students in the discussions could be compared by examining the centrality of nodes corresponding to them in the network; as the nodes' centrality measures their relative importance within a network. Moreover, students could be ranked more explicitly in a concentric centrality graph in which the more central/powerful the node is, the closer it is to the center (Figure 3).

Communities: Collaborative Groups

Reffay et al. [27] and Haythornthwaite [15] emphasized the practical importance of identifying cliques and clusters for the understanding of academic collaborations. Detecting communities in the students interaction networks effectively outlines students collaborations. Moreover, the instructor could easily monitor how these interactions and collaborations are changing over time (Figure 4). Here we used FastModularity [7] as the community detection algorithm, however it could be any other community mining approach.

3.2 Interpreting Term Network

Interpreting the phrase/term network, depicts the terms used in the discussions and the relations between these terms. While finding the hierarchical communities in this network summarizes the debated topics. For word sense community detection, it was observed that different sets of terms used to discuss different topics [4]. Hence the communities in the term network speak for the topics exchanged in the discussions. Furthermore, choosing each of these topics would

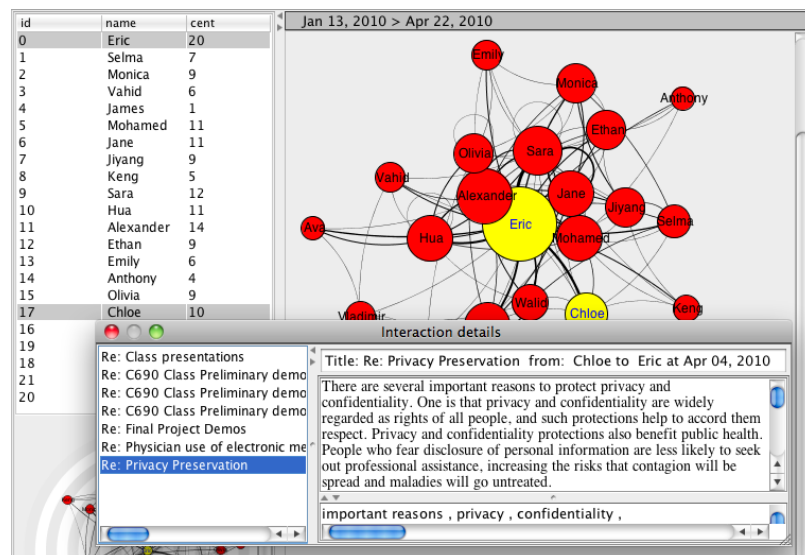


Figure 2: Visualization of messages in an interaction: the interaction window shows the messages passed between nodes incident to the selected edge: Chloe and Eric. Selecting each message from the left panel would show its title, sender, receiver and content.

outline the students who participated in that topic and the extent of their participation. In the term network, nodes represent noun phrases occurring in the discussions; and edges show the co-occurrence of these terms in the same sentence. Each co-occurrence edge contains the messages in which its incident terms co-occurred; and is weighted by the number of sentences in which these terms co-occurred.

Network Extraction

For building this network, we need to first extract the noun phrases from the discussions, then build the network by setting the extracted phrases as nodes and checking their co-occurrence in all the sentences of every message in order to create the edges.

We have used the OpenNlp toolbox [25] for extracting noun phrases out of discussions. OpenNlp is a set of natural language processing tools for performing sentence detection, tokenization, pos-tagging, chunking, parsing, and etc. Using sentence detector in OpenNlp, we first segmented the content of messages into sentences. The tokenizer was used to break down those sentences to words. Having the tokenized words, we used the Part-Of-Speech tagger to determine their grammatical tags – whether they are noun, verbs, adjective, etc. Then using the chunker, we grouped these words into phrases, and picked the detected noun phrases, which are sequences of words surrounding at least one noun and functioning as a single unit in the syntax.

For obtaining better sets of terms to represent the content of the discussions, pruning on the extracted noun phrases was necessary. We removed all the stop words, and split the phrases that have stop word(s) within into two different phrases. For example the phrase “privacy and confidentiality” is split into two terms: “privacy”, and “confidentiality”. To avoid having duplicates, the first characters were converted to lower case (if the other characters of the phrase are in lowercase) and plurals to singular forms (if the singular form appeared in the content). For instance “Patients”

would be “patients” then “patient”. As a final modification, we removed all the noun phrases that just occurred once.

Communities: Discussion Topics

We could represent the important terms in any type of textual data using Tag Clouds (see Figure 7). However in Meerkat-ED we also provided a hierarchy of topics discussed. For this purpose, the term network is further analyzed to group the terms co-occurring mostly together. These groups represent the different topics discussed in the messages and could be obtained by detecting the communities in the term network. This idea is similar to work done in [4]. For creating the hierarchy of the topics, we applied a community mining algorithm repeatedly to divide one of the current connected components of the network, until the size of all components is smaller than a threshold, or the division of any of the components would result in a loose partitioning.

Participation Range

Based on the detected term communities, the participation of students and how wide their participation are could be validated. In other words, students who participated in different topics could be considered more active than students that just talked about a smaller number of topics. This evaluation could be examined by selecting each student and checking how many topics he/she participated in.

4. CASE STUDY

In this section, we illustrate the practical application of Meerkat-ED on our own case study data. Here, Meerkat-ED is used for visualizing, monitoring and evaluating participation of students in a discussion forum. The data set we have used is obtained from a postsecondary course. The course titled Electronic Health Record and Data Analysis, was offered in Winter 2010 at the University of Alberta. The permission to use the anonymized course data for research purposes was obtained from all the students regis-

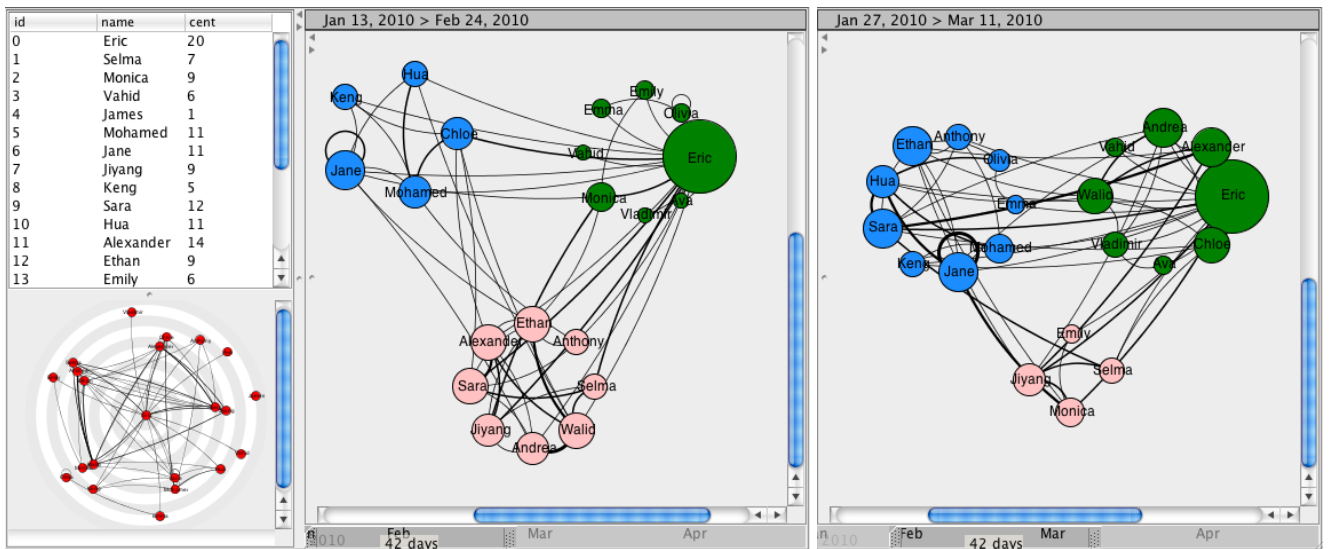


Figure 4: Collaborative Student Groups: these graphs illustrate change in the detected collaborative groups over time, where different groups are depicted in different colours.

tered in the course, at the end of the semester so as not to bias the communications taking place. This data is further anonymized by assigning fake names to students and replacing any occurrence of first, last or user name of the students in the data (including content of the messages in discussion forums) with the assigned fake name. We also removed all email addresses from the data.

In the chosen course, as is also usual in other courses, the instructor initiated different discussion threads. For each thread he posted a question or provided some information and asked students to discuss the issue. Consequently students posted subsequent messages in the thread, responding to the original question or to the response of other students. This course was offered using Moodle which is a widely-used course management system. Moodle like other CSCL tools, enables interaction and collaborative construction of content, mostly using its Forum tool which is a place for students to share their ideas [20]. Only using Moodle, to evaluate student participation the instructor is limited to shallow means such as the number of posts per thread and eventually the apparent size of messages. The instructor would have to manually monitor the content of each interaction to measure the extent of individual participation, which is hard, time consuming and even unrealistic in large classes or forums with large volume, where different participants can be assigned to moderate different discussions and threads.

To assess participation, we build and analyze two kinds of networks from these information: the social network of students and the network of the terms used by them. The instructor of the course denoted the usefulness of the results of these analysis in evaluating the participation of students in the course. Like in [28] where the authors noted that using SNA it was easy to identify the “workers and the lurkers” in the class, in this case study, the instructor reported that using Meerkat-ED it was easy to have an overview of the whole participation and it was possible to identify influential students in each thread as well as identify quiet students or unvoiced opinions, something that would have been impos-

sible with the simple statistics provided by Moodle. More importantly, focusing on the relationships in the graph one can identify the real conduit for information rather than simply basing assessment of participation on message size or frequency of submissions. Learners who place centrally in the network as conduit for the information control and can cause more knowledge exchange which is desirable in an online class. Regardless of the frequency of messages, their size or content, if they do not have influence, their authors remain marginal and sit on the periphery of the network (See Figure 3). This role of conduit of information versus marginal students can change during the course of the semester or from one discussed thread to the other. The systematic analysis of centrality of participants per topic discussed provided by Meerkat-ED allowed a better assessment of the participation of learners at each discussion topic level.

4.1 Interpreting Students Interaction Network

As explained before, first of all we have to extract the students network from the discussion thread. Figure 1 shows the visualized network of students in the course. The size of the nodes corresponds to their degree centrality in the network – the number of incident edges. This means that the bigger a node is, the more messages the student represented by that node sent and received.

The thickness of the edges in the network represents the weight of interactions which is based on the number of messages in the interaction of communicating students. Choosing an edge would bring up a pop up window that shows these messages as illustrated in Figure 2.

The next step is to analysis the leadership of the students based on their centrality in this network. The nodes’ centrality is depicted by the size of the nodes in the visualized network as illustrated in Figure 1. Moreover, students could be ranked more explicitly in a concentric centrality graph in which the more central/powerful the node is, the closer it is to the center, as presented in Figure 3.

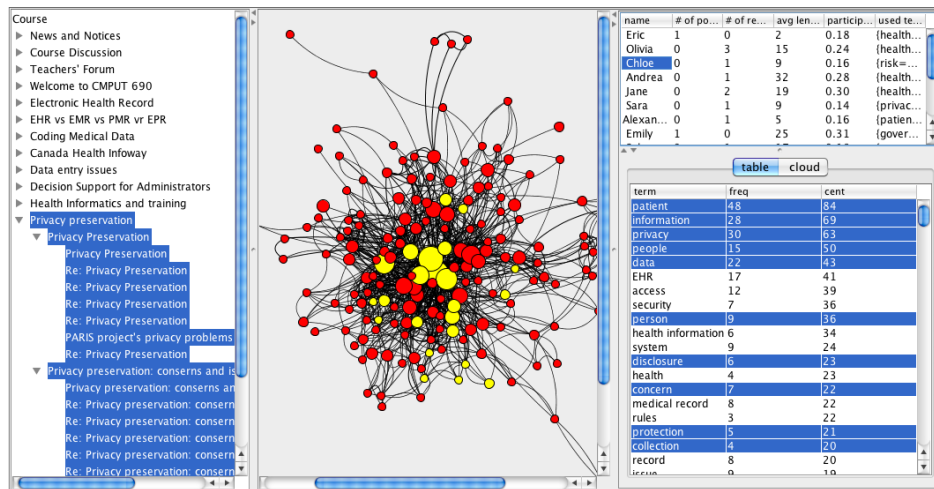


Figure 5: Visualized Term Network: The left panel lists the discussion threads in the course. The middle panel shows the network of terms in the selected set of discussions. The upper right panel shows list of students that participated in the selected discussions, along with some statistics about their participation such as number of posts, replies, etc. The bottom right panel shows the terms used in these discussions and highlighted on the network (middle). Selecting each student, would outline the terms used by that student.

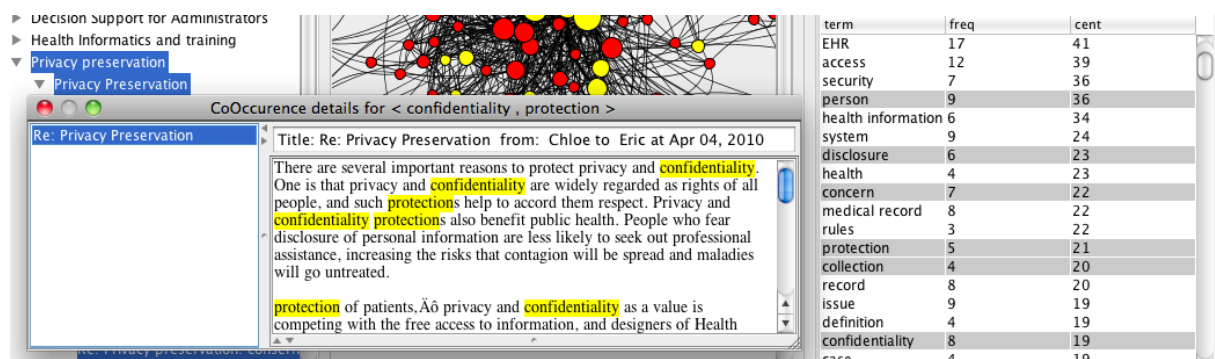


Figure 6: Co-occurrence of terms: selecting a co-occurrence edge would bring up a pop up window that shows the messages these incident terms co-occurred in, highlighting the corresponding terms in the content.

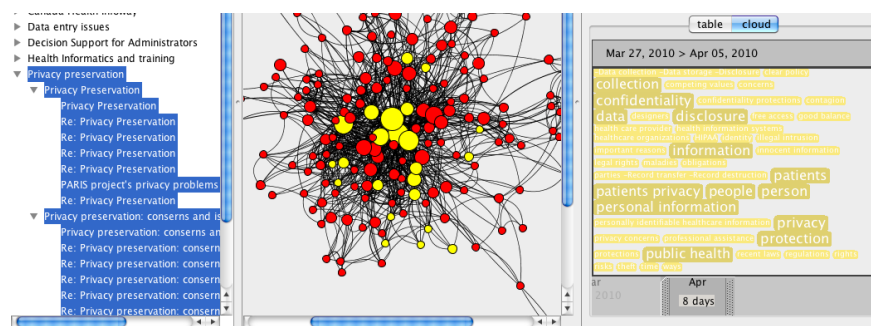


Figure 7: TimeCloud: the right panel shows dynamic cloud of important terms used in these selected discussions for a selected time window.

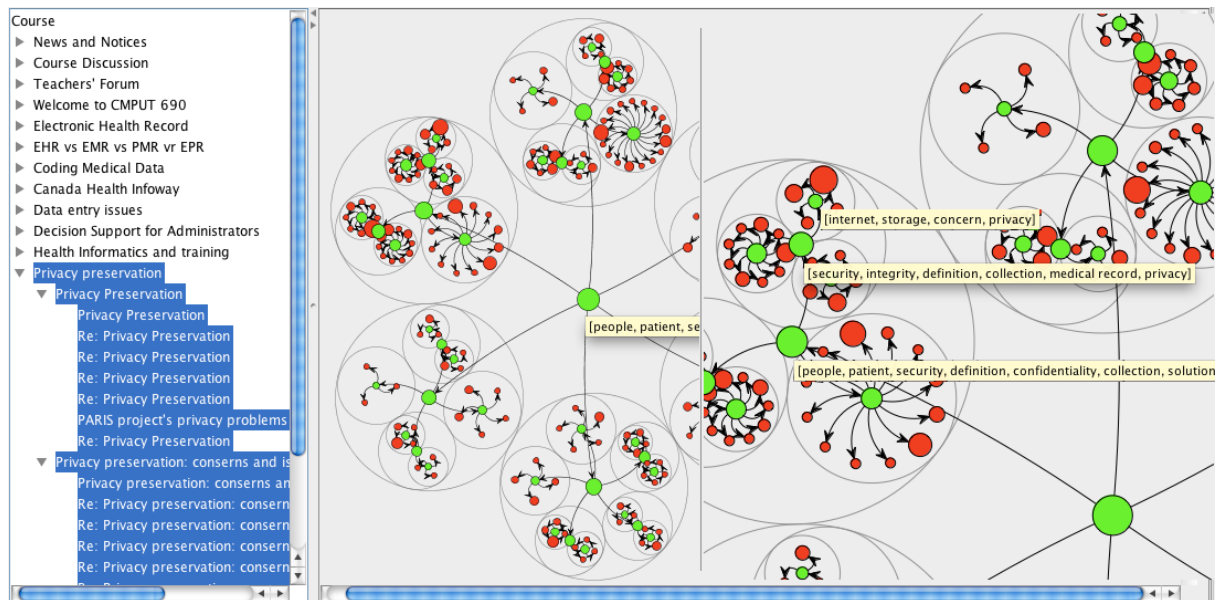


Figure 8: Term communities (Hierarchy of Topics): The gray circles outline the communities boundaries and the green nodes represent the community representatives. Each community representative is accompanied with its top 10 phrases in its community. These could be seen in the tooltip in the figure. The Left panel is zoomed to provide a closer look. Here, the topic is roughly about "privacy, security and confidentiality of medical records".

4.2 Interpreting Term Network

For this specific course, we extract the term network from the discussion forum. Figure 5 presents the visualization of this term network, where the size of the nodes represents the frequency of their corresponding terms and the thickness of edges represents the weight of the co-occurrences (i.e. the number of sentences in which incident terms occurred together). Selecting an edge would show these messages as illustrated in Figure 6.

In this visualization the instructor would see a list of the discussion threads in the course while selecting any set of those discussions/messages would bring up the corresponding term network, along with the list of terms occurring in them and the list of students that participated in these selected set of discussions/messages. Selecting any of these terms would show the students that used that term. Likewise, selecting any of the students would outline the terms used by that student, as illustrated in Figure 5; which is highlighting the terms discussed by the student named Chloe. The difference between the number of terms discussed by the students could help the instructor to compare the participations of the students: students who discuss more terms participate more as well.

In order to further analyze the term Network, as explained before, we group the terms co-occurring mostly together. Figure 8 shows the detected topics (term communities) in the network given in Figure 5. The green nodes show the representative nodes of communities. Each representative node, contains 10 most central terms of the terms in the community it represents. The size of the representative nodes corresponds to the number of terms in their communities; while the size of the leaf nodes, terms, is related to their frequency, same as the term network. Similar to the term network, here also one could select a set of terms, usually

within a topic, to see who participated in a discussion with that topic and to what extent, as illustrated in Figure 9.

5. CONCLUSIONS

In this paper we elaborated on the importance of social network analysis for mining structural data and its applicability in the domain of education. We introduced social network analysis and community mining for studying the structure in relational data. We illustrated the place and need for social network analysis in the study of the interaction of users in e-learning environments; then summarized some recent studies in this area.

We also proposed Meerkat-ED, a specific, practical and interactive toolbox for analyzing students' interactions in asynchronous discussion forums. For any selected period of time, our toolbox prepares and visualizes overall snapshots of participants in the discussion forums, their interactions, the leaders/peripheral students, and collaborative student groups. Moreover, it creates a hierarchical summarization of the discussed topics, which gives the instructor a quick view of what is debated online. It further illustrates individual student participation in these topics, measured by their centrality in the discussions on that topic, their number of posts, replies, and the portion of terms used by them. We believe exploiting the mining abilities of this toolbox would facilitate fairer evaluation of students' participation in online courses.

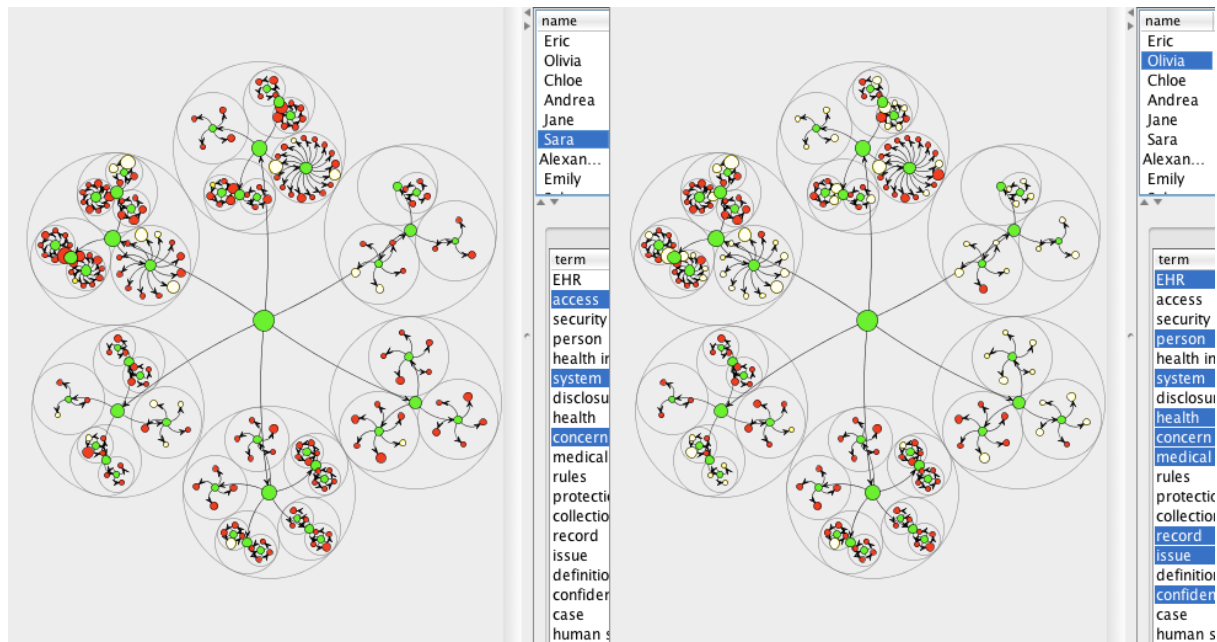


Figure 9: Range of participation in different topics: selecting each topic, would outline the students who participated in a discussion with the topic, and the terms in that topic. Here we could see that Olivia has participated in a wider range of topics compared to Sara. Vice versa, students who participated in each topic and their contribution could be seen in the upper right panel by selecting each topic.

6. REFERENCES

- [1] Blackboard. http://en.wikipedia.org/wiki/Blackboard_Learning_System.
- [2] A. Calvani, A. Fini, M. Molino, and M. Ranieri. Visualizing and monitoring effective interactions in online collaborative groups. *British Journal of Educational Technology*, 2009.
- [3] J. Chen, J. Fagnan, R. Goebel, R. Rabbany, F. Sangi, M. Takaffoli, E. Verbeek, and O. Zaiane. Meerkat: Community mining with dynamic social networks. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 1377–1380, dec. 2010.
- [4] J. Chen, O. R. Zaiane, and R. Goebel. An unsupervised approach to cluster web search results based on word sense communities. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, pages 725–729, Washington, DC, USA, 2008. IEEE Computer Society.
- [5] J. Chen, O. R. Zaiane, and R. Goebel. Detecting communities in large networks by iterative local expansion. In *CASoN*, pages 105–112, 2009.
- [6] H. Cho, G. Gay, B. Davidson, and A. Ingrassia. Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education*, 49(2):309–329, 2007.
- [7] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111, 2004.
- [8] T. Daradoumis, A. Martínez-Monés, and F. Xhafa. A layered framework for evaluating on-line collaborative learning interactions. *Int. J. Hum.-Comput. Stud.*, 64(7):622–635, 2006.
- [9] S. Dawson. Seeing the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5):736–752, 2010.
- [10] S. Dawson, A. Bakharia, and E. Heathcote. Snapp : Realising the affordances of real-time sna within networked learning environments. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. De Laat, D. McConnell, and T. Ryberg, editors, *Proceedings of the 7th International Conference on Networked Learning*, pages 125–133, 2010.
- [11] M. de Laat, V. Lally, L. Lipponen, and R.-J. Simons. Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1):87–103, March 2007.
- [12] Erlin, N. Yusof, and A. A. Rahman. Students’ interactions in online asynchronous discussion forum: A social network analysis. *Education Technology and Computer, International Conference on*, 0:25–29, 2009.
- [13] A. Gruz. *Automated Discovery of Social Networks in Online Learning Communities*. PhD thesis, University of Illinois at Urbana-Champaign, June 2009.
- [14] A. Gruz and C. A. Haythornthwaite. The analysis of online communities using interactive content-based so-

- cial networks. extended abstract. In *Proceedings of the American Society for Information Science and Technology (ASIS&T) Conference*, pages 523–527, Columbus, OH, USA, October 2008.
- [15] C. Haythornthwaite. *Building social networks via computer networks: Creating and sustaining distributed learning communities*, pages 159–190. Cambridge University Press, 2002.
 - [16] S. Hrastinski. What is online learner participation? a literature review. *Computers & Education*, 51(4):1755–1765, December 2008.
 - [17] A. Laghos and P. Zaphiris. Sociology of student-centred e-learning communities: A network analysis. In *IADIS international conference*, Dublin, Ireland, July 2006. e-Society.
 - [18] A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80:056117, Nov 2009.
 - [19] A. Marin and B. Wellman. *Handbook of Social Network Analysis*. Sage, forthcoming, 2010.
 - [20] Moodle. <http://en.wikipedia.org/wiki/Moodle>.
 - [21] NetMiner. <http://www.netminer.com/NetMiner/>.
 - [22] M. E. J. Newman. Detecting community structure in networks. *Eur. Phys. J.B*, 38:321–330, 2004.
 - [23] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69, 2004.
 - [24] K. Nurmela, E. Lehtinen, and T. Palonen. Evaluating CSDL log files by social network analysis. *Computer Support for Collaborative Learning*, 1999.
 - [25] OpenNlp. <http://opennlp.sourceforge.net/README.html>.
 - [26] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.
 - [27] C. Reffay and T. Chanier. How social network analysis can help to measure cohesion in collaborative distance-learning. In *Computer Supported Collaborative Learning*, pages p. 343–352. Kluwer Academic Publishers, May 2003.
 - [28] B. Sundararajan. Emergence of the most knowledgeable other (mko): Social network analysis of chat and bulletin board conversations in a CSDL system. *Electronic Journal of e-Learning*, 8:191–208, 2010.
 - [29] UCINET. <http://www.analytictech.com/ucinet/>.
 - [30] WebCT. <http://en.wikipedia.org/wiki/WebCT>.
 - [31] P. A. Willging. Using social network analysis techniques to examine online interactions. *US-China Education Review*, 2(9):46–56, 2005.