RCEE & RHEd2010
Kuching,Sarawak
7 – 9 June 2010

Automated Social Network Analysis for Online Discussion

Azizah Abdul Rahman [a], **Erlin** [b], **Norazah Yusof** [c]

[a] Department of Information Systems, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Johor Bahru, 81310, Malaysia
[b] Department of Information Technology, STMIK-AMIK Riau, Pekanbaru, 28294, Indonesia
[c]Department of Software Engineering, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Johor Bahru, 81310, Malaysia

Abstract

Social network analysis (SNA) is often employed by educators to analyze online discussion in web based learning to serve various purposes regarding the communication structure of the group and other structural characteristics of a discussion forum. However, current learning management system provide limited information for analyzing interactions among participants. Usually, educators use SNA independently and construct adjacency matrix manually before being analyzed. Therefore, this paper proposes a tool to generate automated social network analysis that is embedded into Moodle as the learning management systems. This system allows educators to analyze online interaction dynamically and simultaneously for the purpose of activity assessment. The tool provides visual representation of the result in map and graphical form. This paper also describes the use of network indicators of social network analysis to assess the level of participation and communication structures through online discussion. Degree centrality, density, map and graph theory were applied to quantitatively define the network interaction. The finding shows that automated social network is useful in simplifying the analysis process efficiently and comprehensibly.
Keywords: automated social network, online discussion, adjacency matrix, centrality, density, graph, map

1. Introduction

Online discussions have gained significant popularity and provide an opportunity to study the characteristic of communication structures of the group. From the perspective of social network analysis, communication structure may be put through an objective analysis process to infer the outcome of the communication. In a teaching and learning context, the communication structure able to reflect the learning process, the learners' characteristics and the learning outcome.

Understanding the nature of relationships and connections between participants is the building block of social network analysis. Wasserman & Faust (1994) stated that social network analysis (SNA) is the study of the structure of social interactions. In educational setting, social network maps and graphs are employed to represent the structure of interactions/relationships among participants. It's based on the tenet that the structure of communication contains information about the level of participation, identifying who are central actors and other structural characteristics of the online discussion.

Social network analysis is often employed in formal social, economic, educational research work and other fields. The peculiarity of this perspective is that it focuses not only on individuals or other social units, but also on the relationship between them.

SNA is an active research area, and its application in learning management system (LMS) particularly in online discussion analysis is challenging. Moodle as the most known LMS generically provide online discussion for interaction among participants and able to give information on the number of messages that posted by participants (Alvaro & Joanne, 2007). Unfortunately, it unable to provide a comprehensive evaluation on the level of interaction in online discussions. The data obtained from the collected interaction is analyzed manually. However, it is very tedious and likely to make mistake.

This paper proposes a tool to analyze communication structures of online discussion automatically and visualized them in graph form and map that educators easily understood. This system is expected to be beneficial to the educators as well as the researchers in assessing the students' participation in online discussion. Besides that, employing automated social network analysis in routine day-to-day teaching would become more practical and viable.

2. Background

Studying the social network analysis (SNA) of online interaction can serve several of purposes. In general, the purpose of social network analysis is to understand the communication structures of online discussion in e-learning environment. Previous research, study on SNA to identify students' interactions in online discussion. For Example, de Laat, et al., (2007) investigated the pattern of interaction in networked learning and computer supported collaborative learning; and Pena-Shaff & Nicholls (2004) studied the students' interaction and meaning construction in computer bulletin board discussion. Another purpose of social network analysis is to assist in online learning discussion where the quantity of interaction is considered prominent (Suh & Lee, 2006).

Social network analysis has also been applied in monitoring the learning progress of students. For example, Willging (2005); Erlin, et al., (2009) applied centrality and density to identify who are central actors in the network and the level of engagement among participants, and from it inferred the learning progress of the cohort.

Unfortunately, most of them used SNA independently. Currently, LMS does not provided information regarding the communication structures since it is not embedded with SNA. The educators of an online class are not supported with structural indicators that would allow them to evaluate the participation and interaction among students in their classes. In many instances, only statistical information, such as frequency of postings is provided. However, it is not a very useful measurement for the interaction activity.

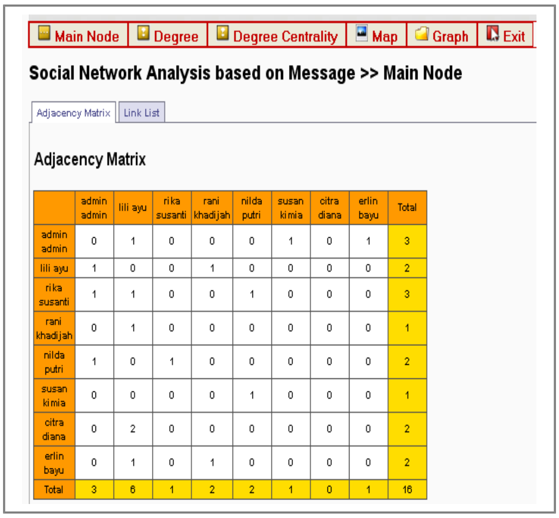
Traditionally, educators or researchers calculate the quantity of interaction manually and then convert them in adjacency matrix before being analyzed by SNA. Furthermore, educator will re-build a new adjacency matrix for each different session from the same online interaction. It is labour-intensive task, time consuming and low accuracy. Therefore, this study proposes an automated SNA tool that can be embedded into LMS in order to provide a dynamic picture of online interaction. The following section will describe how the automated tool work and show the automation changes occur in every session.

3. Creating Adjacency Matrix

In the first step, the tool will create an adjacency matrix. It depicts the current state of the activity in online interaction. The educators can learn and monitor to detect the less active student at early stage and encourage them to participate actively in the next session.

Table 1,2 and 3 shows three sessions of the online discussion in the adjacency matrix form. The data presented in this matrix is captured automatically from online discussion. The adjacency matrix dynamically adjusts to the new data according to any changes post and reply that have been made by each participant. This is as basis for further processing by the tool to determine the network indicators of SNA.

Table 1. Adjacency Matrix of Session 1



Main Node | Degree | Degree Centrality | Map | Graph | Exit

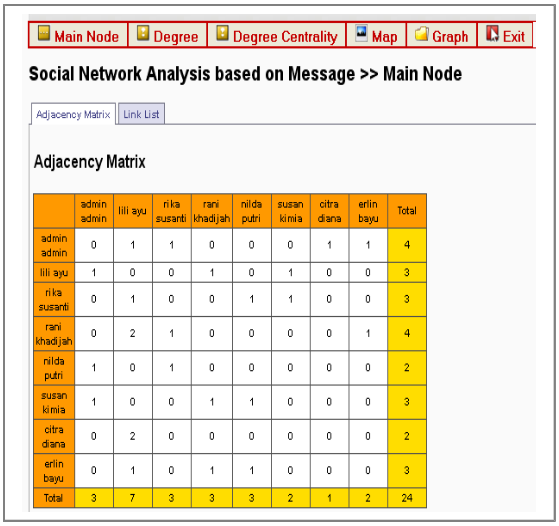
Social Network Analysis based on Message >> Main Node

Adjacency Matrix | Link List

Adjacency Matrix

	admin admin	lili ayu	rika susanti	rani khadijah	nilda putri	susan kimia	citra diana	erlin bayu	Total
admin admin	0	1	0	0	0	1	0	1	3
lili ayu	1	0	0	1	0	0	0	0	2
rika susanti	1	1	0	0	1	0	0	0	3
rani khadijah	0	1	0	0	0	0	0	0	1
nilda putri	1	0	1	0	0	0	0	0	2
susan kimia	0	0	0	0	1	0	0	0	1
citra diana	0	2	0	0	0	0	0	0	2
erlin bayu	0	1	0	1	0	0	0	0	2
Total	3	6	1	2	2	1	0	1	16

Table 2. Adjacency Matrix of Session 2



Main Node | Degree | Degree Centrality | Map | Graph | Exit

Social Network Analysis based on Message >> Main Node

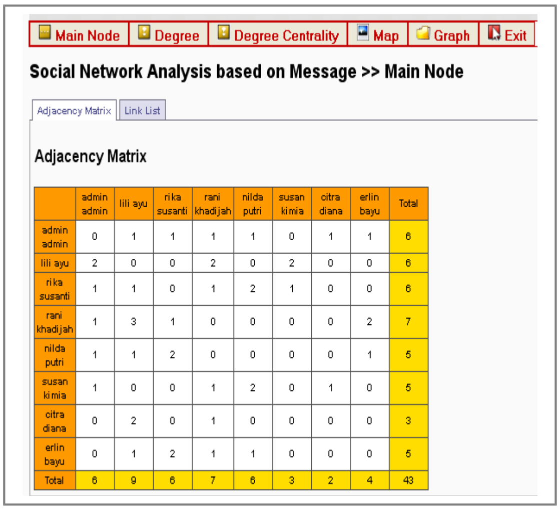
Adjacency Matrix | Link List

Adjacency Matrix

	admin admin	lili ayu	rika susanti	rani khadijah	nilda putri	susan kimia	citra diana	erlin bayu	Total
admin admin	0	1	1	0	0	0	1	1	4
lili ayu	1	0	0	1	0	1	0	0	3
rika susanti	0	1	0	0	1	1	0	0	3
rani khadijah	0	2	1	0	0	0	0	1	4
nilda putri	1	0	1	0	0	0	0	0	2
susan kimia	1	0	0	1	1	0	0	0	3
citra diana	0	2	0	0	0	0	0	0	2
erlin bayu	0	1	0	1	1	0	0	0	3
Total	3	7	3	3	3	2	1	2	24

Table 1 show the total number of interactions that occurs in the first session is 16. In the second session that is shown in table 2, total number of interactions increased to 24. Admin, student Lili and student Erlin make one additional interaction while student Susan and student Rani, make two and three additional interaction respectively. Other students make no changes. Moreover, in the third session (table 3), a total of interaction becomes 43 which is almost double the number of interaction in session 2. Meaning that student are more active in posting their ideas or suggestions in the last session.

Table 3. Adjacency Matrix of Session 3



Main Node | Degree | Degree Centrality | Map | Graph | Exit

Social Network Analysis based on Message >> Main Node

Adjacency Matrix | Link List

Adjacency Matrix

	admin admin	lili ayu	rika susanti	rani khadijah	nilda putri	susan kimia	citra diana	erlin bayu	Total
admin admin	0	1	1	1	1	0	1	1	6
lili ayu	2	0	0	2	0	2	0	0	6
rika susanti	1	1	0	1	2	1	0	0	6
rani khadijah	1	3	1	0	0	0	0	2	7
nilda putri	1	1	2	0	0	0	0	1	5
susan kimia	1	0	0	1	2	0	1	0	5
citra diana	0	2	0	1	0	0	0	0	3
erlin bayu	0	1	2	1	1	0	0	0	5
Total	6	9	6	7	6	3	2	4	43

It is clearly illustrated the temporal of online class discussion at any session and also showed which participants responded and posts to each other and how often they did so.

4. Creating Current Network

In order to measure the current network the tool will calculate four network indicators of SNA. There are degree, degree centrality, density and network degree centralization index. Saltz, et al., (2004) argued that the simplest and most intuitive, within a student-focused online discussion, is degree centrality. This is the measure of interaction regardless of the send/receive directionality (i.e. it measures the volume of activity/messages). Degree centrality is presented by in-degree centrality-counting the number of replies to the messages posted by the student and out-degree centrality-counting the number of messages sent by the student. The formulas of degree centrality included in-degree centrality and out-degree centrality are as follows:

$$\left(di(Mi) = \frac{di}{g-1} \right) \quad (1)$$

$$\left(do(Mo) = \frac{do}{g-1} \right) \quad (2)$$

In (1), $d_i(M_i)$ is a participant's in-degree centrality, d_i is the sum of messages received by the participant from other participants. In (2), $d_O(M_O)$ is the participant's out-degree centrality, d_O is the sum of messages that the participant sends toward others and g is the number of participants in the group.

Secondly, a density that provides a measure of the overall connections between the participants. This gives an indication of the level of engagement in the network (Scott, 2000). The formula to calculate density can be expressed as:

$$\left(Density = \frac{W}{n(n-1)} \right) \quad (3)$$

In (3), W is the number of communicative link and n is the maximum number of possible link.

Lastly, network degree centralization index is an indicator for analyzing the network as a whole based on actor's degree centrality. It gives an idea about the dependency of the network on the activity of small group of participants (Martinez et al., 2003). More precisely, network degree centralization index is defined as

Sum over node *a* of (max centr. In G – centr. of node *a*)

Max value of the sum in all graphs on the same # of node

All the above formulas are used as the basis for making automated algorithm in network density, degree centrality and network degree centralization index of the online discussion.

Table 4, 5 and 6 shows the output summary from three network indicators; network density, in-degree and out-degree. This picture is derived from the adjacency matrix tables that have been generated in the previous process. From session 1, 2 and 3 we can see and assess the degree and the network density, which occurs in the online class discussion dynamically.

Table 4. Network Density, In-degree and Out-Degree at Session 1

Output Summary

Network Density : 0.286

Measures	Values	
	In-Degree	Out-Degree
Sum	16	16
Mean	2	2
Std. Dev.	1.73	0.71
Min	0	1
Max	6	3
# of isolate		0

In-Degree and Out-Degree

#	Participant Name	In-Degree	Out-Degree
1	admin admin	3	3
2	lili ayu	6	2
3	rika susanti	1	3
4	rani khadijah	2	1
5	nilda putri	2	2
6	susan kimia	1	1
7	citra diana	0	2
8	erlin bayu	1	2
	Total	16	16

From report on table 4, 5, and 6, it is clear that the tool can give a dynamic illustration of each degree value changes that appears during a certain period of discussion forums. In session 1 (table 4) student Lili received the most reply with the value of 6 (in-degree), while student Citra received the least (she did not get any replies, in-degree is 0). For temporal state, in the session 1, the highest participation rate in posting the message are admin and student Rika with values of out-degree centrality of 3 and the lowest is student Rani with 1 as the values of out-degree.

It looks different in the second session (table 5). Student Rani who was previously a student of the lowest participation rate (out-degree is 1), in the second session, occupied the highest position with the admin (out-degree is 4). In the last session (table 6) student Rani achieved the first position in the number of posting message with the value of out-degree 7. Student Lili was at top position levels in receiving replies from others with value of out-degree 9.

Table 5. Network Density, In-degree and Out-Degree at Session 2

Output Summary

Network Density : 0.429

Measures	Values	
	In-Degree	Out-Degree
Sum	24	24
Mean	3	3
Std. Dev.	1.66	0.71
Min	1	2
Max	7	4
# of isolate		0

In-Degree and Out-Degree

#	Participant Name	In-Degree	Out-Degree
1	admin admin	3	4
2	lili ayu	7	3
3	rika susanti	3	3
4	rani khadijah	3	4
5	nilda putri	3	2
6	susan kimia	2	3
7	citra diana	1	2
8	erlin bayu	2	3
	Total	24	24

Table 4, 5 and 6 also shows the network density values that indicate how active the participants in the online discussion and can show how dense is the overall participation. The network density values increased from one session to the next session. In session 1, the value of network density is 0.286 which means that few participants connected to other participants. In the second session density values have increased to 0.429 while the third session network density value is 0.768 which means that more than 50% participant connected to each others. When the density is 0, the network is without any connection; and when the density is 100% or 1, all the participants of a network are connected to one another. Based on the network density value in the three tables (0.286, 0.429, 0.768), it shows that the participation become more active at every next session. Thus, educators can then analyze the interaction pattern of the group during discussion.

Table 6. Network Density, In-degree and Out-Degree at Session 3

Output Summary

Network Density : 0.768

Measures	Values	
	In-Degree	Out-Degree
Sum	43	43
Mean	5.38	5.38
Std. Dev.	2.12	0.84
Min	2	3
Max	9	7
# of isolate		0

In-Degree and Out-Degree

#	Participant Name	In-Degree	Out-Degree
1	admin admin	6	6
2	lili ayu	9	6
3	rika susanti	6	6
4	rani khadijah	7	7
5	nilda putri	6	5
6	susan kimia	3	5
7	citra diana	2	3
8	erlin bayu	4	5
	Total	43	43

Table 4, 5 and 6 also shows the network density values that indicate how active the participants in the online discussion and can show how dense is the overall participation. The network density values increased from one session to the next session. In session 1, the value of network density is 0.286 which means that few participants connected to other participants. In the second session density values have increased to 0.429 while the third session network density value is 0.768 which means that more than 50% participant connected to each others. When the density is 0, the network is without any connection; and when the density is 100% or 1, all the participants of a network are connected to one another. Based on the network density value in the three tables (0.286, 0.429, 0.768), it shows that the participation become more active at every next session. Thus, educators can then analyze the interaction pattern of the group during discussion.

Table 7 shows the output summary of the degree centrality and network degree centralization index in the last session. Participant in-degree centrality varied between 0.2857 and 1.2857 and out-degree centrality varied between 0.4286 and 1. It shows that all participants play even role in sending messages, while some participants receiving more messages than the others.

Table 7. Degree Centrality and Network Degree Centralization Index at Session 3

Output Summary

Measures	Values	
	In-Degree Centrality	Out-Degree Centrality
Mean	0.7678	0.7679
Std. Dev.	0.3234	0.1697
Min	0.2857	0.4286
Max	1.2857	1

In-Degree and Out-Degree Centrality

#	Participant Name	In-Degree Centrality	Out-Degree Centrality
1	admin admin	0.8571	0.8571
2	lili ayu	1.2857	0.8571
3	rika susanti	0.8571	0.8571
4	rani khadijah	1	1
5	nilda putri	0.8571	0.7143
6	susan kimia	0.4286	0.7143
7	citra diana	0.2857	0.4286
8	erlin bayu	0.5714	0.7143
	Total	6.1427	6.1428

Network Degree Centralization Index

In : 59.18 %

Out : 26.53 %

This table also shows the network degree centralization score for in-degree is 59.18% means that few students who have score high in-degree centrality (student Lili and student Rani). The high score shows the uneven interaction in receiving the messages. Most of the student score is low in in-degree centrality (student Citra, student Susan and student Erlin) and the other three are medium (Admin, Student Nilda and student Rika). Different conditions shown in the network degree centralization for the out-degree centrality is 26.53%. This value is low, means that overall participants is even in sending the messages, not only dominated by few participants. Moong (2007) argued that when the measure of network degree centralization is large, it means that few participants are highly central and the remaining participants occupy much less central positions in the network. Conversely, if the network centralization is low, it means that the network is populated. Compare these two values, it shows that these groups involved more better in sending messages.

5. Map and Graph

As a way to address the visualization of online discussion, we presented graph in figure 1-3 and map in figure 4-6 for marking up the rhetorical structure of online interaction in threaded forum. The SNA tool enable automatically analysis level of participation and visualization them in graph and map.

This graph is set up using 5 layers. The limitation value of each layer depends on the maximum and minimum value of degree centrality. The formula for the difference in the value of max and min values can be express as:

$$D = (N\max - N\min)/5 \quad (3)$$

List 1 is drawing algorithm for graphical of degree centrality.

---

a. Layer detection: identify the number of layers; L=1…5
b. Boundary detection: identify the boundary value of each layer; L5 = Min of degree centrality, L4 = L5 + *D*,…, L1=L2+*D*
c. Display: arrange the nodes in a clockwise
d. Permutation of node: for each layer find the permutation of each node according to its degree centrality value

---

List 1. Graph Algorithm

The placement of each student that represent as a node in the layer depends on its degree centrality. The greater the degree centrality will be closer to the center of the circle. Conversely, the smaller the degree centrality will be far away from the center of the circle (located at the outermost layer).

In the first session student Rani and student Susan in the position farthest from the center of the circle (at fifth layer) which means that they have the lowest degree centrality value (out-degree=1). Striking changes seen in the second session where student Rani at the position closer to the center of the circle which means that she is the most active student in posting a message to another student. Different circumstances presented by the student Citra, which at the end of the whole session he was in the lowest position with out-degree value of the lowest, so that her position was at fifth layer.

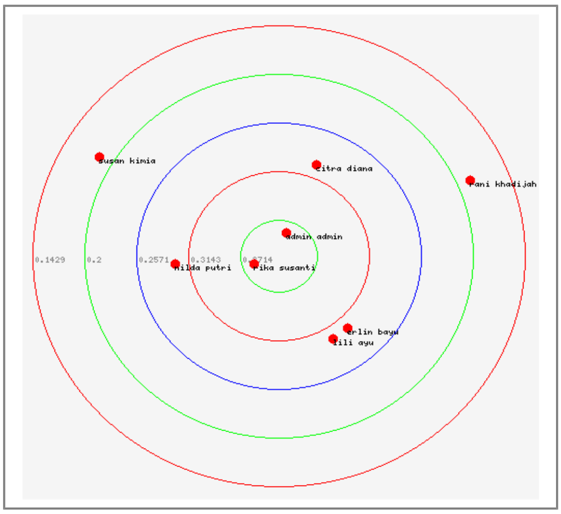


Figure 1. Graph of Out-Degree Centrality at Session 1

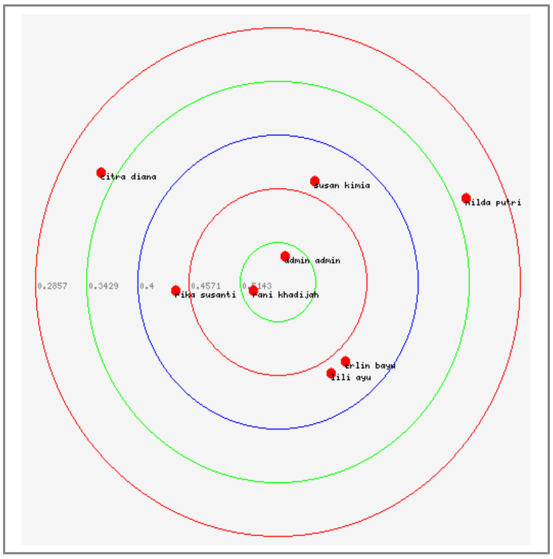


Figure 2. Graph of Out-Degree Centrality at Session 2

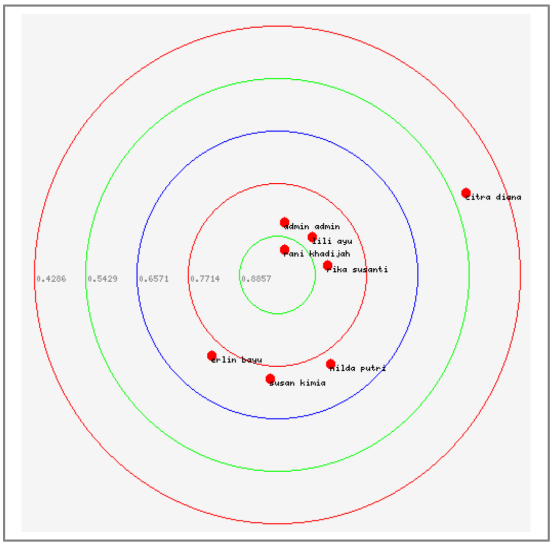


Figure 3. Graph of Out-Degree Centrality at Session 3

Moreover, there is vast literature and research in concerning automatic map drawing. Maps are represented by symbols such as circle or boxes for every node and drawing an arc between two nodes if they are connected by an edge. If the edge is directed, the direction is indicated by drawing an arrow.

According to Cruz (2009), there are five general styles of maps drawing; polyline drawing, planar straight-line drawing, orthogonal drawing, planar orthogonal straight-line drawing and visibility representation. Some drawings are better than others in conveying information on the map. Aesthetic criteria attempt to characterize readability by means of general optimization goals. This research employed planar straight-line drawing in angular resolution.

Algorithm drawing map

Finding a realization of a map $M= (V, E)$ where $V=\{v_1, v_2, v_3, v_4\}$ and $E= \{\{v_1, v_2\}, \{v_3, v_4\},\{v_1, v_4\}, \{v_1, v_3\},\{v_3, v_2\}, \{v_2, v_4\}\}$ and given an abstract planar graph $M= (V, E)$ find a map drawing of it using straight-line segments for edges. Two edge segments may intersect in (at most) one common endpoint.

Based on the algorithm above, a map was created for analyzing link interaction as shown in figure 4, 5 and 6. The graph below presents the link relations among students at session 1,2 and 3 of discussion. The map shows that, as time goes on, the scope of linkage expands from one student to others, and that the degree among students intensifies. The map provides a meaningful analysis on factors such as who is active in online discussion, who is central participant, whom the participant interacted with and who the leader of interaction.

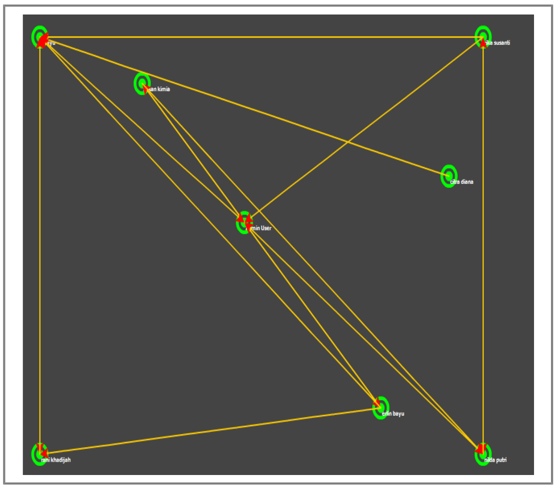


Figure 4. Map of Out-Degree Centrality at Session 1

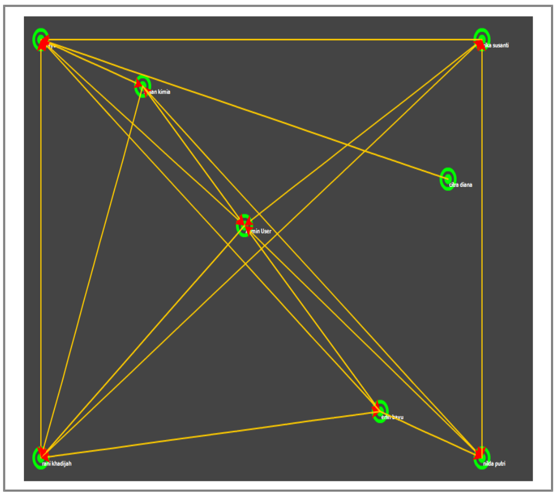


Figure 5. Map of Out-Degree Centrality at Session 2

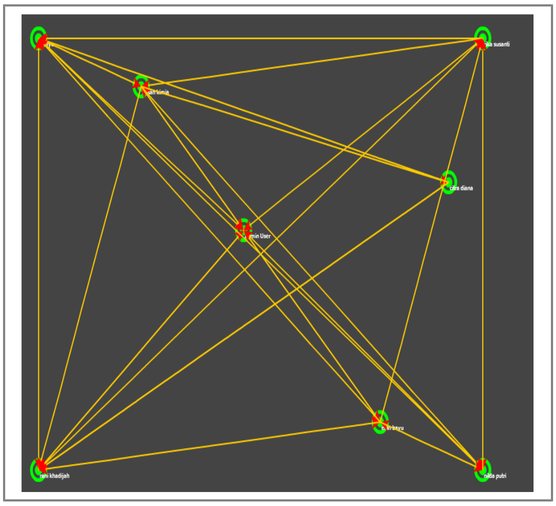


Figure 6. Map of Out-Degree Centrality at Session 3

Fig. 4, 5 and 6 illustrates the interaction patterns between eight students in a directed graph. From figure 4 it can be said that student Lili get many replies and highest in term of in-degree centrality while Admin and student Rika highest in out-degree centrality meaning that they active in posting a message to other student. Student Rani, student Susan and student Citra interact only with student Lili. Moreover, student Citra interact in one way, meaning that they had been isolate from others. Also from this graph educators can detect who is involved with the discussion, who is active participant and who is lurker.

5. Conclusion

This research proposed an automated tool for social network analysis that embedded into LMS. It is a possible solution for automatically depicting and analyzing relations that are established between participants in online interactions. It provides the visualization of online discussion in graph form and map presentation. The tool is found to be helpful for educators to analyze interaction dynamically. It enables educators to monitor students' interaction.

Acknowledgements

This research was supported by Ministry of National Education of the Republic of Indonesia and Universiti Teknologi Malaysia.

References

1. S. Wasserman & K. Faust, Social Network Analysis, Cambridge University Press, 1994.

2. R.F. Alvaro & B.L. Joanne, Interaction visualization in web-based learning using iGraph, Proceedings of the eighteenth conference on Hypertext and Hypermedia, Manchester, United Kingdom, 2007, pp. 45-46
3. M. de Laat, V. Lally, L. Lipponen, & R.J. Simon, Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis, Journal of Computer Supported Collaborative Learning, *2*(1), 2007, pp. 87-103.
4. J.B. Pena-Shaff, & C. Nicholls, Analyzing student interactions and meaning construction in computer bulletin board discussions, Journal of Computers & Education, 42(3), 2004, pp. 243-265.
5. H.J. Suh & S.W. Lee, Collaborative learning agent for promoting group interaction, ETRI Journal, 28(4), 2006, pp. 461-474.
6. P.A.Willging, Using social network analysis technique to examine online interaction. US-China Education Review, 2(9), 2005, pp. 46-56.
7. Erlin, Y. Norazah, & A.R. Azizah, Students' interaction in online asynchronous discussion forum: A social network analysis, International Conference on Education Technology and Computer, Singapore, 2009, pp. 25-29.
8. J.S. Saltz, S. R. Hiltz, & M. Turoff, Student social graphs: visualizing a student's online social network, Proceeding of Computer Supported Cooperative Work, Chicago, Illinois, USA, 2004, pp. 596-599.
9. J. Scott, Social Network Analysis: A Handbook, 2nd ed, London, Sage, 2000.
10. A. Martinez, Y. Dimitriadis, B. Rubia, E. Gomez & P. de la Fuente, Combining qualitative evaluation and social network analysis for the study of classroom social interactions, Journal of Computers & Education, 41 (4), 2003, pp. 353-368.
11. M. K. Soong, Equicentrality and network centralization: A micro–macro linkage, Journal of Social Networks 29, 2007, pp. 585-601.
12. I.F. Cruz, Graph drawing tutorial, 2009, available: http://www.graphdrawing.org/literature/gd-constraints.pdf