Journal of Engineering and Applied Sciences 12 (Special Issue 5): 6974-6978, 2017
ISSN: 1816-949X
© Medwell Journals, 2017

Influential Students Detection in an e-Learning Discussion Forum Using Degree Centrality Metric

Satrio Baskoro Yudhoatmojo, Rizky Andika and Harry Budi Santoso
Faculty of Computer Science, Universitas Indonesia, Kampus UI,
16424 Depok, Jawa Barat, Indonesia

Abstract: Learning environment has been extended not just in physical environment but also in virtual environment. Active learning activity such as discussion can take place virtual environment using discussion forum. Learning management system provides this feature and several other learning activities. Students use the discussion forum to discuss issues or enquire information of related topics stimulated by the teacher or lecturer or by the students themselves. Interaction between students would take place in this discussion forum, a student may post in-reply to others or add a new post. The layout of a typical discussion forum is leading to the difficulty of analyze the interaction network between students and thus identifying who is the most influential student or the most dominant student in the discussion forum. In our research, we used theories from social network analysis (a part of complex network) to build an interaction network between students in the forum and used the degree centrality metric to detect the most influential student in the discussion forum. We extracted the online discussion forum data from the database of a moodle-based learning management system used in our university which is called SCeLE, short for student-centered learning environment. We focused on one course which has several parallel classes. Despite having a very low number of students participating in the online discussion forum, we have managed to identify the most influential student in the discussion forum using degree centrality metric.

Key words: e-Learning, online discussion forum, degree centrality, social network analysis, complex network, metric

INTRODUCTION

Learning environment has been greatly enriched by the use of technology. Technology used in learning environment does not only help the teacher or lecturer to transfer their knowledge to the students in physical learning environment but also in virtual environment. With the help of internet connection and Learning Management System (LMS) application, the course materials can be accessed at anytime and anywhere. Discussions between lecturer and students or among the students about the course topic can, not only, take place in class and/or virtual environment such as online discussion forum provided by the LMS. Lecturer can provide students some stimulation/trigger questions to start the discussions. The students themselves can start the discussions. Thus, the students would experience an active learning as stated by Maybee *et al.* (2016) which is beneficial to student learning and engagement.

Online discussion forum in an LMS is a good place for students to exchange knowledge regarding the topic being discussed and as mentioned by Costley (2016) to develop both private reflection and higher order discourse. Students also learn to actively articulate their thoughts into words. This is very important, especially, for those students who are having a difficult time in expressing their thoughts in class.

In our university, we adapt a moodle-based learning management system (https://moodle org) which we called Student-Centered Learning Environment (SCeLE for short) (Willging, 2005). SCeLE has a feature for online discussion. In our research, we focused on one course for examining its discussion forums. For the purpose of this research, we requested the discussion forum data from the system owner which is Kantor Sumber Daya Pembelajaran (Center for Learning Resources) Universitas Indonesia. We generated the discussion interaction network from the requested data. From the network data, we analyzed it to detect the most influential student in each forum using methods from social network analysis.

Literature review: In this study, we reviewed several theories and works which are related to our work. Social network is a set of social agents such as people,

Corresponding Author: Satrio Baskoro Yudhoatmojo, Faculty of Computer Science, Universitas Indonesia, Kampus UI, 16424 Depok, Jawa Barat, Indonesia

organizations, etc., which may have relations or links to one another. According to Rabbany *et al.* (2011), there are six common characteristics in a network structure: the most influential social agents, individual agent with the most outgoing relation to other agents, the most incoming relation from other agents and the one that does not do anything, how one agent related to other agents, agents who are frequently communicated to each other, agent who acts as an information broker in the network and comparison of possibility of relation in the network. Moreover, according to Wasserman and Galaskiewicz (1994) social influence is a strategic place for researching about social network. Social influence relates social structure relation with the behavior of the social agents who form the network. One of the social network analysis metric that can be used for analyzing the network for social influence is degree centrality. According to Elizabeth degree centrality is a metric for calculating the sum of direct edges for a node. We can use this to adjust the size of node in the visualization to depict the importance of that node as mentioned by Rabbany *et al.* (2011).

One of the previous work which are related to our work is done by Rabbany *et al.* (2011). They developed a toolbox called Meerkat-ED for visualizing overall snapshots of discussion forum's participants their interactions and the leader/peripheral students by applying SNA methods. They used online discussion forum data of electronic health record and data analysis course from University of Alberta, Canada in 2010. They did two things in their research: interpreting student interaction network using Meerkat-ED and interpreting term network using OpenNlp toolbox. Similar work was also carried out by Borgatti (2005). He used discussion forum data from a course in Blackboard called online learning management system. He formatted the data into adjacency matrix. He used UCINET Software for analyzing the interaction network and MAGE Software to create an interactive visualization and 3D visualization of the interaction network.

MATERIALS AND METHODS

Our research data were from a course in our university which is called MPKT (Mata Kuliah Pengembangan Karakter Terintegrasi/Integrated Character Development Course). Originally there are 77 distinctive classes but we omit 18 of them because there is only one participant in the forum discussion thus it does not count as a discussion (Silva and Saraiva, 2015). The total number of students are 2,393 students but there are only 278 students who are participated in the online discussion and by omitting forum discussion having only one student, we finally have only 262 students as our data. These students are undergraduate students on their first year.

In term of methodology in our research we adapt the work of Juliana. The study was focusing on biological interaction networks. Even though, the study focused on specific domain, we believe that we can adapt this work on another domain. In Juliana there are four steps in applying social network analysis metrics:

- Step 1: mapping data types and interactions
- Step 2: defining key-questions and variables
- Step 3: choosing appropriate SNA metrics
- Step 4: performing biological analysis with SNA

RESULTS AND DISCUSSION

Step 1 (mapping data types and interactions): In this step, there is an initial activity which comprises the exploratory task of which data types are available. Our research used online discussion forum data of a course in our university's LMS. Currently, there is no specific student interaction network data of the course's online discussion forum data. Thus, we have to extract data from the database of the LMS and reformat data in a way that usable for the computational tool that we are going to use in Step 4. We retrieved the raw data by requesting the raw data from the administrator of the LMS. The administrator queried the data we need from the LMS and the result is a database table with the following structure:

```
Create table c (
        username VARCHAR(100),
        idnumber VARCHAR(255),
        firstname VARCHAR(100),
        lastnname VARCHAR(100),
        id DOUBLE,
        parent DOUBLE,
        created DOUBLE,
        fullname VARCHAR(254),
        subject VARCHAR(255),
        message BLOB
)
```

The raw data that we got were in the form of SQL. We stored the raw data in PostgreSQL DBMS. The next activity is to extract and clean the raw data to get data that we need to form the student interaction network. Our aim is to create a student interaction network data in the format used by Pajek Software. As mentioned by Mrvar and Batagelj (2016) Pajek is software package for analysis and visualization of large networks. This format is can also be used by complex network library called NetworkX which we used in our analysis. In order to build the student interaction network, we need the list of students as the nodes or vertices and the data about who replies to whom as the links or edges. We transferred the raw data in the database to Microsoft Excel. The two important columns the id and the parent columns. These two columns are the post identification number, more specifically the parent represents the previous post replied by the post with identification number of id.

The links or edges of the network are formed by pairing the id and parent into a tuple. Since, the raw data consisted of several classes of the same course, we need to separate them into their classes. The separation is done by looking at the subject column which represents the class. The result of this separation is having 77 different classes and each class is saved into a text file. We analyzed the result and decided omit classes that had less than two vertices. The reason behind this is that to have an interaction between students there need to be at least two students (two vertices) participating in the interaction, thus if there is only one vertices then there is no interaction with other students. This omission led into 59 different classes.

Finally, we need put the vertices and edges into the Pajek format. We wrote two small Java programs to process each class into Pajek format. The first program took the text file as an input and process it to extract the vertices from the list of edges and return an output file with specific structure as seen from the output sample:

```
*Vertices
0 “0"
33298  “33298”
33706  “33706”
34695  “34695”
36240  “36240”

*Edges
33988  0
33706  33298
34695  33706
36240  34695
```

The second program used the output of the first program as an input can perfecting it into the correct Pajek format. The output file of the second program can be seen on the output sample:

```
*vertices 8
0  “0”
1  “33321”
2  “33335”
3  “33349”
4  “33643”
5  “33644”
6  “33656”
7  “33783”
*edges
1_0
2_1
3_1
4  2
5  3
6  1
7  1
```

At this point, we have not grouped the id and parent to their respective student. The grouping process was done manually. For students who have more than one post id, we promoted one of the id as the single identifier for that student and replaced the other id for the same student with that single identifier. The result of this process is having smaller number of vertices. By the end of this process, we have the student interaction network data in the format that can be used for analyzing using computation tools that we used.

Step 2 (defining key-questions and variables): In the second step, we defined the key-questions that we want to answer from the proposed problem and chose the variables that can be used for analysis. The proposed problem in our research came from the limitation of our university’s LMS, that is, there is no feature for analyzing student interaction network in discussion forum to identify which students are influential students (or as mentioned in Juliana the leader and peripheral students). We derived the key-question from the proposed problem and that is: “Who is the most influential student in each discussion forum?”. The variables that we need for this analysis are list of students, post information and post reply information. The list of students is going to be the nodes, the post information is used to identify which student is posting data to the forum and post reply information are going to be the source for creating edges between nodes.

Step 3 (choosing appropriate sna metrics): In this step, we chose the appropriate SNA metrics for analyzing the student interaction network in the online discussion forum. According to Daly and Haahr (2007) to select the most appropriate metrics, we must consider which aspect of the interaction network is to be analyzed: the network structure or the role of each node in the network. In our research, we chose the role of each node in the network aspect because of our aim in this research is to identify the most influential student in each discussion forum. Thus, we are interested in the role of the students in the network (the leader/influential or peripheral).

The SNA metric that is suitable for the purpose of our analysis is the centrality metric. The centrality metric itself has different types of centrality such as degree centrality, betweeness centrality, closeness

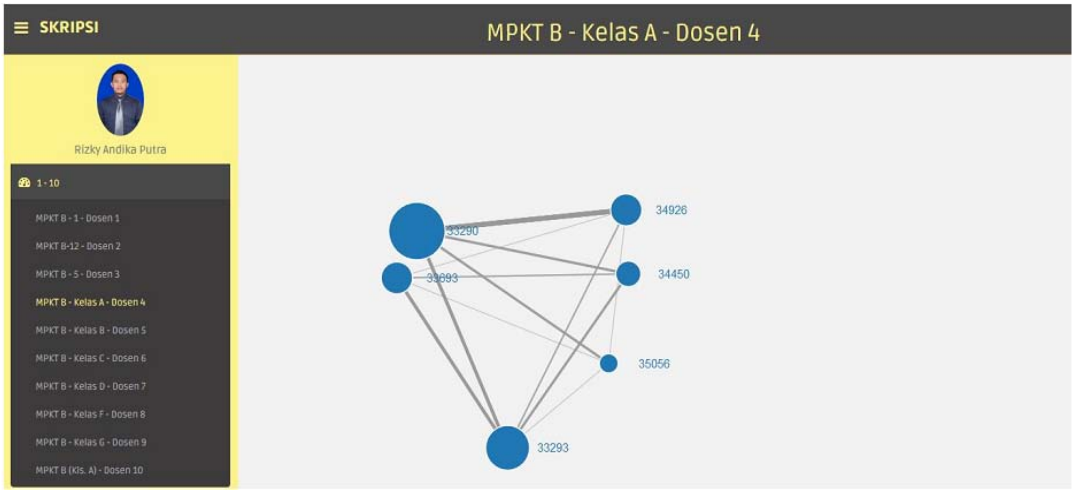
≡ SKRIPSI
MPKT B - Kelas A - Dosen 4
Rizky Andika Putra
1 - 10
MPKT B - 1 - Dosen 1
MPKT B-12 - Dosen 2
MPKT B - 5 - Dosen 3
MPKT B - Kelas A - Dosen 4
MPKT B - Kelas B - Dosen 5
MPKT B - Kelas C - Dosen 6
MPKT B - Kelas D - Dosen 7
MPKT B - Kelas F - Dosen 8
MPKT B - Kelas G - Dosen 9
MPKT B (Kls. A) - Dosen 10
34926
33290
33693
34450
35056
33293

Fig. 1: Student interaction network visualization

centrality and eigenvector centrality. In our research, we chose to use degree centrality because it suited for the goal of our work. Degree centrality is the number of ties incident upon a node. Stephen also mentioned that degree centrality can be regarded as a measure of immediate influence (the ability to infect others directly or in one-time period). This is exactly what we need in our work. Thus, degree centrality is appropriate to be used in our work.

Step 4 (performing analysis with SNA): In our research, the calculation of degree centrality of our student interaction network was done with the help of a library called networkX. NetworkX is a python-based library for the creation, manipulation and the study of the structure, dynamics and functions of complex networks. Our dataset has been categorized based on each course class which is taught by different lecturer. For each of the dataset, we apply degree centrality method from networkX to calculate the degree centrality of the student interaction network of that dataset. The following is the example of the application of that method:

```
file =nx. read_pajek(“output MPKT B – Dosen 1. paj”)
doc = nx. degree_centrality(file)
```

In the code snippet above, the file is a container for the result of reading a pajek format file. Then, we apply the method to that file. The result of applying that method can be seen:

```
{
     “33293”: 2.8000000000000003,
     “33290”: 3.6,
     “35056”: 1.2000000000000002,
     “34926”: 2.0
     “34450”: 1.6,
     “33693”: 2.0
}
```

For each node, the method gave its degree centrality value. The node that has the highest values can be considered as the most influential node or the leader of that network. Additionally, we also used the number of edges method from networkX for illustrating the interaction frequency between two nodes (students).

It is quite difficult to grasp the idea of the most influential node or the leader and peripheral nodes and interaction frequency with just by looking at the result of those methods. Another contribution that we made in this research is that we visualized that idea by showing the student interaction network, made the size of the nodes respectable to their degree centrality value and made the thickness of the edges based on the number of edges between the two nodes. In our work, the visualization was done using a JavaScript library called d3js. D3js is a JavaScript library for manipulating documents based on data Mike Bostock. In order to integrate the network analysis and the visualization, we used a python based web application framework called Flask (Armin, 2016) because it is a very lean framework and easy to use for our proof of concept. For the web GUI, we used a bootstrap template made by Carlos (2015) called DashGum. An example of the visualization can be seen in Fig. 1.

On Fig. 1, we can see that student with ID 33290 has the highest degree centrality based on the size of the node. We can conclude from the calculation and the visualization that this student is leading the discussion. Furthermore, the thickness of the edges shows the interaction frequency. There is a quite high frequency from other students to the student with ID 33290. We can conclude that student with ID 33290 influenced other students to interact/discuss topics or issues posted by that student.

Our work analyzed 59 classes with the minimum of two nodes to show the interaction network between one student with the other student (s). Within these 59 classes, we were able to derive similar visualization as in Fig. 1 and conclude that the most influential students or the leader can be identified from our work. Nevertheless, our work still has limitation. The first limitation is that our dataset for each class is still very small which shows that not all students in that class participated in the online discussion. The non-participating students are not included in our node list hence the very low number of nodes. The second limitation is that we were only using the degree centrality to identify the most influential student.

CONCLUSION

This research is a preliminary of our work on analyzing online discussion forum of our university's learning management system using social network analysis metrics. This research is able to show that the most influential students can be identify using degree centrality metric. To enable an easier identification, we visualized the result.

LIMITATIONS

The limitation that we have are: the very low number of students (or nodes) in our student interaction network and using only degree centrality metrics in our analysis.

RECOMMENDATIONS

For future work, we would like to gather more data and enriched our analysis using a variety social network analysis metrics. By having more data, we are interested in finding communities within an online discussion forum and analyze the topic discussed within that community.

ACKNOWLEDGEMENTS

We would like to thank Mr. F.A. Triatmoko H.S. from Kantor Sumber Daya Pembelajaran (Center for Learning Resources) Universitas Indonesia who had helped us in providing the dataset of our research.

REFERENCES

Armin, R., 2016. Flask-web development, one drop at a time. The Hip Flask Company, Sheffield, England. http://flask.pocoo.org.

Borgatti, S.P., 2005. Centrality and network flow. Social Netw., 27: 55-71.

Carlos, A., 2015. DashGum free dashboard: Black tie free handsome bootstrap themes. A Tin Can Technologies Company, Providence, Rhode Island. http://blacktie.co/2014/07/dashgum-free-dashboard/.

Costley, J., 2016. The effects of instructor control on critical thinking and social presence: Variations within three online asynchronous learning environments. J. Educ. Online, 13: 109-171.

Daly, E.M. and M. Haahr, 2007. Social network analysis for routing in disconnected delay-tolerant manets. Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, September 9-14, 2007, Montreal, Canada, pp: 32-40.

Maybee, C., T. Doan and M. Flierl, 2016. Information literacy in the active learning classroom. J. Acad. Librarianship, 42: 705-711.

Mrvar, A. and V. Batagelj, 2016. Analysis and visualization of large networks with program package Pajek. Complex Adapt. Syst. Model., 4: 1-8.

Rabbany, R., M. Takaffoli and O.R. Zaiane, 2011. Analyzing participation of students in online courses using social network analysis techniques. Proceedings of the 4th International Conference on Educational Data Mining, July 6-9, 2011, Publisher Partners, Eindhoven, Netherlands,-pp: 21.

Silva, J.S. and A.M. Saraiva, 2015. A methodology for applying social network analysis metrics to biological interaction networks. Proceedings of the IEEE-ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 25-28, 2015, IEEE, Paris, France, pp: 1300-1307.

Wasserman, S. and J. Galaskiewicz, 1994. Advances in Social Network Analysis: Research in the Social and Behavioral Sciences. SAGE Publications, Thousand Oaks, USA., ISBN: 9780803943032, Pages: 300.

Willging, P.A., 2005. Using social network analysis techniques to examine online interactions. US. China Educ. Rev., 2: 46-56.