Computer Engineering and Applications Vol. 6, No. 2, July 2017

Comparative classification of student's academic failure through Social Network Mining and Hierarchical Clustering

Andi Besse Firdausiah Mansur, Norazah Yusof

* *Faculty of Computing and Information Technology Rabigh, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia*
abmansur@kau.edu.sa

ABSTRACT

Student academic failure are caused by several factors such as: family relationship, study time, absence, parent education, travel time and etc. This study observe several factors which are related to student academic failure by calculating the centrality degree between students to find the correlation between failure factors for each students. Furthermore, each student will be measured by measuring the geodesic distance for each factors for hierarchical clustering. The flow betwenness measure and hierarchical clustering show the promising result, where students who has similar factors value are tends to be grouped together in the same cluster. The student with high value of flow betwenness is considered as broker of network and play vital character inside network. The result of study is believed can bring important and useful information toward the student performance analysis for future better education.

Keywords: Students Academic failure, Social network analysis, hierarchical clustering, degree centrality.

1. INTRODUCTION

Student performance monitoring is vital to academic monitoring where the result can be useful for school, parent and teacher. Each factors that contribute to the success or failure of student need to be analyzed deeply to produce better education system in the future. Furthermore, Social network mining is one of techniques that are able to reveal the relationship between people that share similar factors. This study focus analyzing the student data that consist of several personal attribute such as: parent status, parent education, study time, free time, or even travel time. Each student will be compared among each other and reveal its relationship

2. RELATED WORKS

Social network analysis (SNA) or Social Network Mining (SNM) is a technique that usually used to analyze the behavior of people inside network or community. It's due to its characteristic that capable to represent social relationship between human or people [1-2]. The other researcher also used SNA to determine the level of participants, core player or even to identify the structure of social interactions [3].

Furthermore, the other learning tools like Wiki also can be used as medium for working together within the same environment [4]. Wiki able to store the page that has contribution on editing the page of wiki. Wiki has several of social learning activities such as editing, uploading, commenting, or tagging, while other researcher focus on studying the attitude of wiki and interaction of Wiki [5-6].

Social learning network analysis focus on analyzing the e-learning domain as a social network. It focus on finding the behavior of student through the networks model of student communication network[8]. This process is believed able to enhance learning and teaching process[9].

The other researcher focus on analysis the text of student message and perform document text analysis (ATA). Most of the previous research focused on analyzing the content of asynchronous discussion such as discussion forums and messaging. However, they are still lacking in term of flow betwenness to find the core flow of the networks. Therefore, the objective of this paper is to fill in those research gaps by focusing on implementing SNA to analyze the student behavior through the factors that contributing toward failures of student. Additionally, previous researcher observe the ties of network through YouTube video and provide an idea to integrate social learning and social network [10-11]. The enhanced tools for learning as well as identifying the group with similar social entities should come up with a general identity. The relationship between learning and study strategies also play important role on student success. It also mentioned that personality peculiarity and individual difference involve on success story of student[12-15]. Providing learning skills and clustering the student based on their online behavior also contribute on student performance. Socio-technical methods are introduced as well to investigate the social site of user and how to evaluate the social media as a tools for collaborative group writing[16-21].

3. MATERIAL AND METHODS

Student data that acquired from P. Cortez and A. Silva is extracted and analyzed according to the social network characteristic. Around the twenty samples of student data is extracted and given further analysis. The data is filtered and only several attributes remain for analysis as shown in Table 1.

The data in Table1 will be processed further by calculating the geodesic distance among students, then it will continued by measuring the centrality degree and perform hierarchical clustering for the given data, refer to Table 2 for student data with geodesic distance, refer to equation 1.

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(1)$$

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 2. Geodesic distance data for students

	Stu1	Stu2	Stu3	Stu4	Stu5	Stu6	Stu7	Stu8	Stu9	Stu10	Stu11	Stu12	Stu13	Stu14	Stu15	Stu16	Stu17	Stu18	Stu19	Stu20
Stu1	0	7	5	15	4	9	7	8	8	6	3	9	11	5	11	8	10	11	11	10
Stu2	7	0	6	10	4	14	10	15	15	10	5	15	17	11	6	15	16	17	17	16
Stu3	5	6	0	11	4	13	10	12	13	10	4	13	15	8	8	13	14	15	14	15
Stu4	15	10	11	0	11	23	19	23	23	19	13	23	25	19	6	23	25	25	25	25
Stu5	4	4	4	11	0	12	9	12	12	9	2	12	14	8	6	12	14	14	14	14
Stu6	9	14	13	23	12	0	5	4	5	4	11	6	4	6	19	5	4	4	4	5
Stu7	7	10	10	19	9	5	0	6	6	3	7	8	8	6	15	7	7	8	7	7
Stu8	8	15	12	23	12	4	6	0	4	6	10	4	5	5	18	2	3	5	4	2
Stu9	8	15	13	23	12	5	6	4	0	6	10	5	6	6	19	3	4	5	4	4
Stu10	6	10	10	19	9	4	3	6	6	0	7	7	7	5	15	7	7	7	7	7
Stu11	3	5	4	13	2	11	7	10	10	7	0	11	13	7	9	10	12	12	12	12
Stu12	9	15	13	23	12	6	8	4	5	7	11	0	4	5	18	4	5	4	5	5
Stu13	11	17	15	25	14	4	8	5	6	7	13	4	0	6	20	5	3	2	3	5
Stu14	5	11	8	19	8	6	6	5	6	5	7	5	6	0	15	5	6	7	7	7
Stu15	11	6	8	6	6	19	15	18	19	15	9	18	20	15	0	19	20	20	20	20
Stu16	8	15	13	23	12	5	7	2	3	7	10	4	5	5	19	0	3	5	5	3
Stu17	10	16	14	25	14	4	7	3	4	7	12	5	3	6	20	3	0	4	2	2
Stu18	11	17	15	25	14	4	8	5	5	7	12	4	2	7	20	5	4	0	3	5
Stu19	11	17	14	25	14	4	7	4	4	7	12	5	3	7	20	5	2	3	0	4
Stu20	10	16	15	25	14	5	7	2	4	7	12	5	5	7	20	3	2	5	4	0

Its started by building the matrix adjacency of the social network by calculating the geodesic distance for each student as shown in Table 2. After the calculation of matrix adjacency there two main experiments which are conducted in this study: Flow betwenness based on Knoke information network and hierarchical clustering are used to measure the behavior of student inside network.

TABLE 1. Student raw data

	Pstatus	Medu	Fedu	Travel time	Study time	famrel	Free time	Go out	absences
Student1	1	1	1	1	2	4	3	2	10
Student2	1	3	2	1	1	5	5	5	16
Student3	1	2	2	1	1	1	2	2	14
Student4	1	2	2	2	2	3	3	3	25
Student5	1	2	2	2	2	4	3	3	14
Student6	1	4	4	1	2	4	4	4	2
Student7	0	4	2	2	1	5	5	5	6
Student8	1	1	1	1	2	3	3	4	2
Student9	1	2	1	2	1	4	5	1	2
Student10	1	4	4	2	2	4	4	4	6
Student11	1	2	2	2	2	4	4	2	12
Student12	1	1	1	2	4	3	1	2	2
Student13	1	3	3	1	3	4	1	2	0
Student14	1	2	2	1	2	3	1	2	6
Student15	1	1	3	3	2	5	2	4	20
Student16	1	0	1	1	2	3	4	2	2
Student17	1	2	2	1	1	3	3	3	0
Student18	1	3	4	2	3	4	2	2	0
Student19	0	3	3	2	1	4	3	2	0
Student20	1	1	1	1	2	4	4	4	0

4. RESULT AND DISCUSSION

The result of experiment is shown that some factors such as parent education, study time and absence has strong correlation each other that contribute to the failure status of student. Figure 1 show the network of student based on their interaction among each other. As seen in Figure1, relation between students 17-20 to student4 has the busiest network with 25 flow, this indicate that student4 is the most important mediator in this network. It can controlled the flow for most of student inside network.

The flow of betwenness centrality is a measurement for actor inside network to determine which actor that depend to another actor when information exchanged occurred, so we can determine which actor that act as broker or control the whole network. For each binary network that consist of collection of vertices v1….vn, the maximum flow betwenness centrality is cmax. The network flow betwenness centralization can be computed through S(cmax-c(vi)) divided by the possible maximum value.

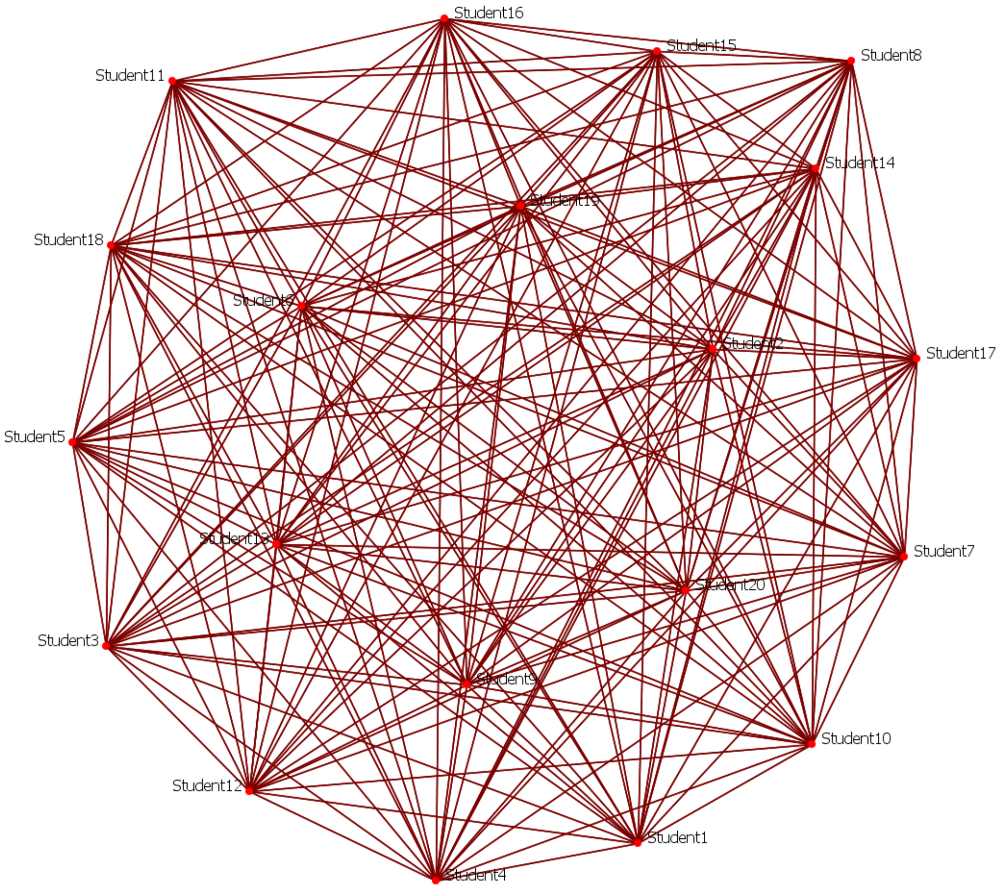

FIGURE 1. Social Network diagram for student data

The c(vi) is known as flow betwenness centrality for vertex vi. The detail of flow betwenness experiment is described in Table 3.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 3. Flow betwenness value for students network

	FlowBet	nFlowBet
Student1	2680.000	4.899
Student2	4408.000	8.141
Student3	3830.000	7.062
Student4	7238.000	13.397
Student5	3586.000	6.600
Student6	2016.000	3.684
Student7	2222.000	4.056
Student8	1844.000	3.352
Student9	2010.000	3.674
Student10	2100.000	3.820
Student11	3078.000	5.650
Student12	2180.000	3.993
Student13	2282.000	4.191
Student14	2032.000	3.684
Student15	5740.000	10.624
Student16	1936.000	3.532
Student17	2034.000	3.723
Student18	2294.000	4.214
Student19	2186.000	4.010
Student20	2110.000	3.871

As described in Table 3, student4 has flow betwenness value 7238 and nflow betwenness 13.397. This value is the highest among network and decided that student 4 can act as broker for this network or the most important person inside this network. So what is the correlation between the failures and the flow betwenness value. As an answer, if Table 1 and Table 3 are compared, it can be found that student 4 has the highest amount of absence from the school (25 absence). Student 4 also has parent with low education level (5^{th} -9^{th} grade). Student 4 also has less study time and more free time or go out time. This mean there are strong correlation between absence, study time and go out time with the failures of student. It also mentioned that low education level of parent also contribute to the failureness of student even though this factor should be studied further.

The second experiment is hierarchical clustering of student network according to their geodesic distance among each other then it continued by classifying the student with similar characteristic into same group. Figure 2, show the result of hierarchical clustering that calculated based on the geodesic distance of student raw data.

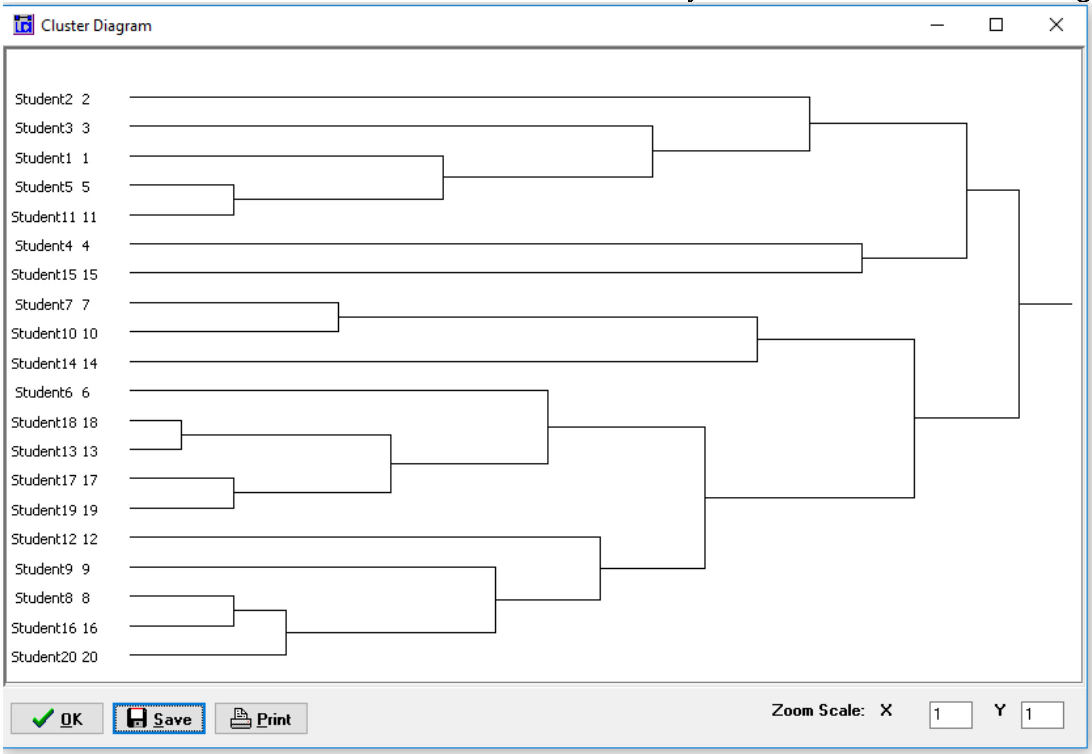

FIGURE 2. Hierarchical clustering result

Figure 2 shows that student who has similar flow betwenness value or similar characteristic tends to be clustered together. Student 4 and 15 has the high flow betweenness value and it's grouped together in the hierarchical clustering.

4. CONCLUSION

Social network analysis is very useful for studying the behavior of user inside network. It can help people to identify the key actor inside network such as a person who are capable on controlling the flow of information inside network. The SNA also beneficial for identifying the cause of failures for student in their study. It is found that high absence rate and less study time has strong contribution towards failures of student. This study has reveal a method and process toward analysis of student failures factors which are very imperative in the direction of better future education system in the school or university.

ACKNOWLEDGEMENTS

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah Saudi Arabia. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

REFERENCES

[1] P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th Future Business Technology Conference, pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-397.

[2] Ji, J., Zhou, C., Wang, Z. & Yan, H. 2011. Maximizing The Community Coverage Of Influence Through A Social Network. AISS: Advances In Information Sciences And

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Service Sciences,Vol. 3, No. 9, Pp. 339 ~ 346, 2011.http://dx.doi.org/10.4156/aiss.vol3.issue9.44
[3] Zhou, B. & Wu, C. 2011. Semantic Model For Social Networking Federation. Aiss: Advances In Information Sciences And Service Sciences, Vol. 3, No. 11, Pp. 213 ~ 223, 2011.
[4] Erlin B., Yusof, N. & Rahman, A. A. 2009. Analyzing Online Asynchronous Discussion Using Content And Social Network Analysis In: Ieee-Ninth International Conference On Intelligent Systems Design And Applications, 2009.http://dx.doi.org/10.1109/isda.2009.40
[5] Ortiza, F. & Fraile, R. 2009. Social Network Featuring Entertainment, Culture And Technology In Spanish Universities: The Infocampus Project The Open Information Systems Journal, 3, 48-53. http://dx.doi.org/10.2174/1874133900903010048
[6] Kepp, S.-J. & Schorr, H. Year. Analyzing Collaborative Learning Activities In Wikis Using Social Network Analysis. In: Chi, Spotlight On Work In Progress~Session 2, April 4-9, Boston, Ma, Usa. Acm, 2009.
[7] Twu, H.-L. 2010. A Predictive Study Of Wiki Interaction:Can Attitude Toward Wiki Predict Wiki Interaction In High-Context Cultures Groups? Journal of Educational Technology Development and Exchange, 3(1), 57-68.
[8] Pasquale De Meo, Fabrizio Messina, Domenico Rosaci, Giuseppe M.L. Sarné, "Combining trust and skills evaluation to form e-Learning classes in online social networks, Information Science,Vol 405,2017,107-122
[9] [R.Hübscher,Assigning students to groups using general and context specific criteria, Learn.Technol. IEEE Trans.3(3) (2010), 178–189.
[10] A.A.Rad,M.Benyoucef, Similarity and ties in social networks: a study of the youtube social network, J.Inf.Syst.Appl.Res.(2014)14–65.
[11] H. Roreger , T.C. Schmidt , Socialize online learning: Why we should integrate learning content management with online social networks, in: Pervasive Computing and Communications Workshops,2012 IEEE International Conference, 2012, pp.685–690.
[12] M. Erdt , A. Fernandez , C. Rensing , Evaluating recommender systems for technology enhanced learning: a quantitative survey, Learn. Technol. IEEE Trans.8(4)(2015)326–
[13] P. Grabowicz, L. Aiello , V. Eguiluz , A. Jaimes , Distinguishing topical and social groups based on common identity and bond theory, in: Proc. of the ACM Int.Conf.WSDM2013,ACM,2013,pp.627–636.
[14] Constantinos M. Kokkinos, Apostolos Kargiotidis, Angelos Markos,The relationship between learning and study strategies and big five personality traits among junior university student teachers,Learning and Individual Differences 43 (2015) 39–47
[15] Komarraju, M., Karau, S.J., Schmeck, R.R., & Avdic, A. (2011). The big five personality traits, learning styles, and academic achievement. Personality and Individual Differences, 51,472-477
[16] Heller, M.L., & Marchant, G.J. (2015). Facilitating self-regulated learning skills and achievement with a strategic content learning approach. Community College Journal of Research and Practice. Advanced online publicationhttp://dx.doi.org/10.1080/10668926.2014.908752.
[17] Yeonjeong Park, JiHyunYu, Il-HyunJo,Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute,Internet and Higher Education 29 (2016) 1–11
[18] Halverson, L. R., Graham, C. R., Spring, K. J., Drysdale, J. S., & Henrie, C. R. (2014). Athematic analysis of the most highly cited scholarship in the first decade of blended learning research. The Internet and Higher Education, 20,20–34.

[19] Wen-Lung Shiau,Yogesh K. Dwivedi,Han Suan Yang,"Co-citation and cluster analyses of extant literature on social networks",International Journal of Information Management 37 (2017) 390–399
[20] Chai, S., & Kim, M. (2012). A socio-technical approach to knowledge contribution behavior: An empirical investigation of social networking sites users, International Journal of Information Management, 32, 118–126
[21] Samuel K.W. Chu, Catherine M. Capio, Jan C.W. van Aalst,Eddie W.L. Cheng,Evaluating the use of a social media tool for collaborative group writing of secondary school students in Hong Kong,Computers & Education 110 (2017) 170-180

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)