Data Analytic for Student Behaviour Classification in Social Learning Network

Andi Besse Firdausiah Mansur[1†], Norazah Yusof[2††], and Nada Omar Bajnaid[3†††]

King Abdulaziz University, Jeddah, Saudi Arabia

Summary
Social Learning network has grown tremendously over past year. This trend has stimulated E-learning system to evolve as social learning network. Meanwhile, the conservative clustering approach on social learning network still less explored due to its association relationship between students are exist. This paper proposed an approach to apply data analytic for revealing the behaviour of student from wiki activities using two major clustering techniques: hierarchical and Minimum Spanning Trees (MST) clustering. The first step is creating the matrix adjacency to determine indegree and outdegree level, and then convert the matrix into geodesic distance matrix. Meanwhile MST created by graph with weighted edge based on the intensity of relationship. The experimental study has shown that student who are possessed the high outdegree value tends to be at the same cluster. The other indication is high outdegree will influence the path in the network. Therefore, it will create shortest route in network, which is indicate the information passed through this student will be broadcasted instantaneously. As shown in the result section student1 and student12 is more active compared to others student. Our approach has been successfully classifying the active and passive student based on their participation on Moodle E-Learning Wiki. This means data analytic using hierarchical clustering and minimum spanning tree can be used to identify student performance or behaviour during study. Consequently, teacher may use it as reference for better e-learning system in the future.

Key words:
Data Analytic, Social Learning Network, Hierarchical Clustering, Minimum spanning Trees, E-Learning.

1. Introduction

Social Network Analysis (SNA) is one of popular technique which is suitable to be used for analysing pattern of user behaviour inside the Social Network (SN). The relationship between users is symbolized using node and edge. Several conservative clustering aaproach are not suitable for social network since it composed from graph. In this paper, we will describe some potential clustering technique which is implemented to analyze the E-Learning data at Universiti Teknologi Malaysia (UTM). The clustering that will be analyzed: Hierarchical clustering and Minimum Spanning Tree (MST). The paper structured is started with introduction of research that mainly discussed the motivation of research. Section 2, focus on discussion for the related research. Section 3 refers to research method and material, whilst section 4 focus on result and discussion. Finally the conclusion of work is presented to summarize the whole result of research.

2. Related Works

There are several conventional clustering technique which is implemented on Social Learning Network(SLN) : Hierarchical Clustering approach by UCINET[1]. Tan.P et.al. mentioned that hierarchical and graph clustering also used frequently on SLN object[2]. The other researcher has come with different solution by using Business System Plan(BSP) to classify the social learning network data. They proposed an approach that can classify every entity inside network into diverse group based on the connection and relationship which is exist in the group [3]. Furthermore the other researcher offer innovative solution for social network clustering by creating new model for clustering through dormant location of the cluster form which can conquer the weakness of Euclidean [4]. They used two assessment: a two-stage maximum likelihood and Bayesian approach using Markov chain Monte Carlo sampling[4]. On the other hand, Mishra et. al., believe that inner solidity and outer slimness for cluster features[5].
There are several classification or clustering approach require central for the cluster (e.g. k-means), this approach will not work against social learning network[2]. Because all data are connected each other that triggering interaction among the node. Thus, an innovative clustering approach is required to solve this research problem. The experiment focused on implementing the agglomerative hierarchical clustering and graph clustering to find the behaviour of student inside E-Learning system at Universiti Teknologi Malaysia. Table 1 show the characteristic of clustering behaviour that will become a key to classify the student during clustering process.

Manuscript received October 5, 2018
Manuscript revised October 20, 2018

Table 1: Clustering features criteria

Features	***Clustering Criteria***
Outdegree	Number of outgoing edge from certain node Represent the number of information that broadcasted to another students
Indegree	Number of ingoing edge for certain node Represent the number of information that read/received by students

Research Hypothesis: Our preliminary hypothesis stated that active student will be likely spread the information that they gain during learning process. Hence, The information that we have recorded on this learning process most likely has strong relationship with theory or material on the university subject. This also mean student who had high outdegree is more knowledgeable compare to the others students since they become centre of information (refer to Figure 1).

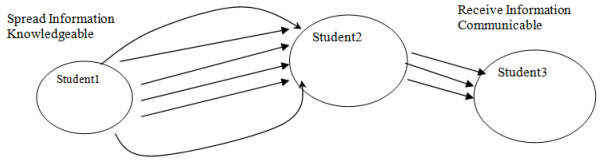

Fig. 1 Illustration of Outdegree and Indegree relationship as features of clustering process

The conservative approach like K-Means or Fuzzy C-Means is not enough for clustering network data like social network due to the large number of dataset and not covering the behaviour inside the network [6-11]. The other approach of clustering is based on collective wisdom that proposed by Agarwal et. al. They proposed the link based clustering (Wis Clus) by grouping the similar labels into new label [6]. Furthermore, other research has focused on student behaviour classification using socio-technical methods to analyse social media and its website to be used as collaborative media for group writing and discussion [12-16]. While other researcher also studied Role-Sphere Influence and Flow betwenness centrality to classify failure rate among students[17].

3. Methodology and Experiment

The data for analysis is obtained from the course code SSC2213-01 and the course name is "Instrumentation in Analytical Chemistry" and the name of the topic is "The topic for Assignment2". The course is handled using Moodle E-Learning system at our university for batch 2010/2011. There are two main parts of the analysis. First is the data collection of the Moodle Wiki data log from E-Learning server. Then it followed by a procedure which is focused on how long the capturing process to obtain the data. Afterward, the data is analyzed by creating the adjacency matrix and followed by calculating the geodesic distance of the matrix adjacency. Based on this distance, hierarchical clustering is created. We have used single link, complete link and WTD_average for the computation. For the graph clustering, we used Minimum Spanning Tree (MST) to cluster the data.

3.1 Data Collection

Data for this research were the log activities of Moodle Wiki on the course code SSC2213-01 that include the updating, editing and creating activities of Wiki. Based on this data we found that 39 users collaborating on one Wiki topics and made interesting interactive communication through Wiki. There are several activities on Wiki: publish Wiki topic, Editing or Updating, Wiki_View and Wiki_Link. All these activities will be categorized as the degree centrality point of view.

3.2 Experimental method

The log activity on Wiki data is captured from Moodle E-Learning system during particular period. The user involvements are categorized based on their contribution inside e-learning. The data need to be processed further by producing matrix of adjacency to reflect the distance among communal communication of user. In order to analyze the Wiki data, it can be exploited through UCINET tool for creating the adjacency matrix of Wiki. In this research, we arrange the relationship of nodes based on indegree and outdegree. The wiki activities such as wiki-add and wiki-edit are considered as outdegree while wiki-view and wiki-link are measured as indegree. After that, we calculate the adjacency matrix as input, then it will be continued by estimating the geodesic distance between each node in the network. Fig.2 describes the relationship of the nodes in plain graph.

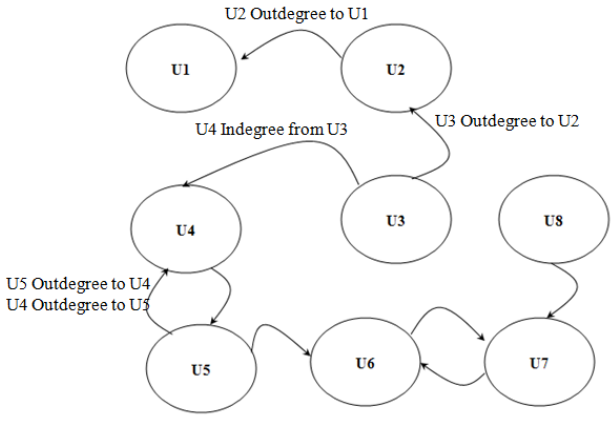

Fig. 2 Plain graph of Social network relationship

Table 2 describe the social relationship between actor/agent in social learning network of E-Learning. Whilst, Table 3 contains the distance of every node to other node which is computed through geodesic computation from matrix adjacency like shown in equation1.

$$\text{GeodesicDistance} = d(i, j)$$
where i=start node, j = destination node (1)

Table 2: Adjacency Matrix for Wiki

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28	U29	U30	U31	U32	U33	U34	U35	U36	U37	U38	U39
U1	0	3	3	3	3	3	4	2	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U12	1	0	1	0	0	1	2	0	0	1	0	0	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
U13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
U15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U21	0	0	1	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1
U22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2: Matrix for Geodesic distance

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_10	S_11	S_12	S_13	S_14	S_15	S_16	S_17	S_18	S_19	S_20	S_21	S_22	S_23	S_24	S_25	S_26	S_27	S_28	S_29	S_30	S_31	S_32	S_33	S_34	S_35	S_36	S_37	S_38	S_39
S_1	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
S_2	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_3	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_5	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_6	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_7	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_8	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_9	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_10	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_11	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_12	1	2	1	2	2	1	1	2	2	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
S_13	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_14	4	4	4	4	1	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4
S_15	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_16	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_17	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_19	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_20	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_21	4	4	1	4	4	4	4	4	1	1	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	1	4	4	4	1	1	1	1	1	1
S_22	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_23	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_24	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_25	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4	4
S_26	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4	4
S_27	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4	4
S_28	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4	4
S_29	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4	4
S_30	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4	4
S_31	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	4
S_32	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4
S_33	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4
S_34	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4
S_35	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4
S_36	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4	4
S_37	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4	4
S_38	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0	4
S_39	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0

4. Result and Discussion

This section will discuss two main data analytics approach which are used to analyze the behaviour of student inside E-learning.

4.1 Hierarchical Clustering on Social Network

As discussed in the previous section that the goal of this paper is to investigate the student behavior through two clustering algorithms in social learning network of E-Learning data. According to Tan et.al. on their book "Introduction to Data Mining", Hierarchical Clustering has three main approaches:

A. Single Link (Minimum Distance)

In this approach, we set the proximity of cluster based on two nearby point in different cluster or in other words the shortest path between two nodes which is placed in different subset (refer to Fig 3) and equation 2.

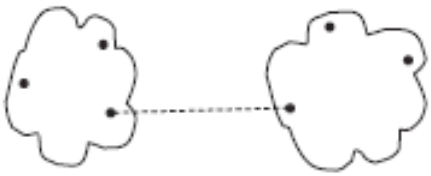

Fig. 3 Single link approach[2]

Minimum or Single Link:

$$\min\{d(a,b): a \in A, b \in B\} \quad (2)$$

B. Complete Link (Maximum Distance)

Complete link is the opposite of Single Link, it use the longest path between two nodes where they in different subset (refer to Fig 4) and equation 3.

Maximum or Complete Link:

$$\max\{d(a,b): a \in A, b \in B\} \quad (3)$$

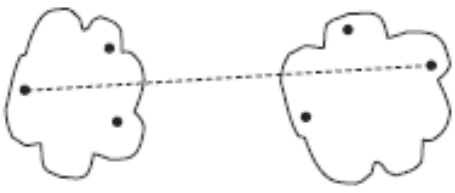

Fig. 4 Complete link approach[2]

C. Group Average

This approach use average pairwise proximities to compute the closeness of the node pairs on different subset(refer to Figure 5) and equation 4.

Group Average

$$\frac{1}{|A||B|}\sum_{a\in A}\sum_{b\in B} d(a,b) \quad (4)$$

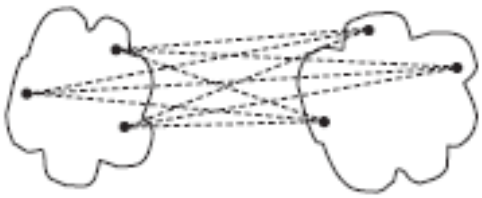

Fig. 5 Group Average approach[2]

As a result of implemented hierarchical clustering to analyze the social network analysis data of E-Learning. The result of clustering is described at Fig 6, 7 and 8.

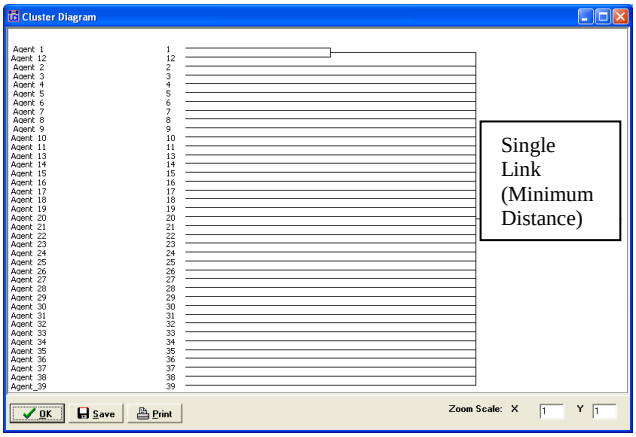

Fig. 6 Hierachical Clustering Using Single Link Approach

Fig.6 is depicting the result of Wiki data clustering based on Matrix adjacency and Geodesic distance at Table 1 and 2. U1(Agent1) and U12(Agent 12) has the highest and biggest outdegree value since this two Matrix very active on supplying the network with information that read by others student. Otherwise, the other student (apart from student 1 and student12) placed in same cluster due to their outdegree value almost the same.

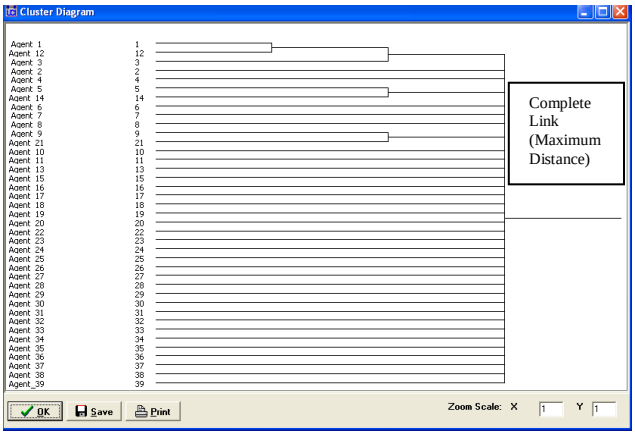

Fig. 7 Hierachical Clustering Using Complete Link

Fig 7 has illustrate the behaviour of student tends to group together when they have similar behaviour. Student 1 and student 12 who have the highest outdegree value is placed in the same cluster. Student5 and student14 has proximity in term of indegree value, this phenomenon also occur on student9 and student21 which is belong to same cluster.

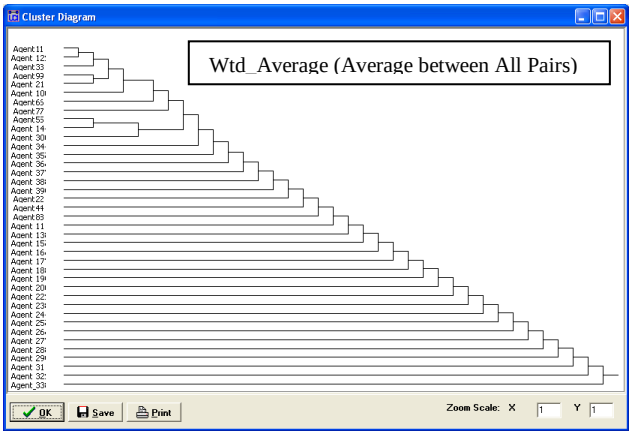

Fig. 8 Hierarchical Clustering Using Group Average

Fig.8 use average between all pairs method to conduct clustering process. Agent1(student1/U1) is grouped with Student12 due to they have high outdegree value. The same treatment also occurred to student5 and student14, student9 and student21. From this characteristic we found that student1 and student12 have the shortest route in network, this mean the information passed through them is very fast. This condition makes student1 and student12 tends to be pro active compared to others student

4.2 Graph Clustering for Social Network (MST Clustering).

In this study we also focus on implementing Kruskal Algorithm to find MST in our network. First, we create a graph network by adding the agent as node and edge between agents is their relationship. The weighted of the edge depend on the outdegree and indegree value.

Kruskal Algorithm (G,ω) **:**

Input: A connected undirected graph $G=(V,E)$ with edge weights ω_e we

Output: A minimum spanning tree defined by the edges X

$\mu \in V$:

$makeset(\mu)$

for all

$X = \{\ \}$

sort the edges E by weight

for all edges $\{\mu,\upsilon\} \in E$, in increasing order of weight:

if find (μ)≠find (υ):

add edge $\{\mu,\upsilon\}$ to X

$union(\mu,\upsilon)$

Fig.9 and 10 demonstrate the application of minimum spanning tree (MST) which shown in two colour, black edge is the original path/route that are not chosen during MST calculation. The purple edge and node colour is the effect of MST calculation, the purple edge mean that node will be drag to be same at same cluster with their pair node. For example, node 28 is tends to be grouped with node12 rather than node1.

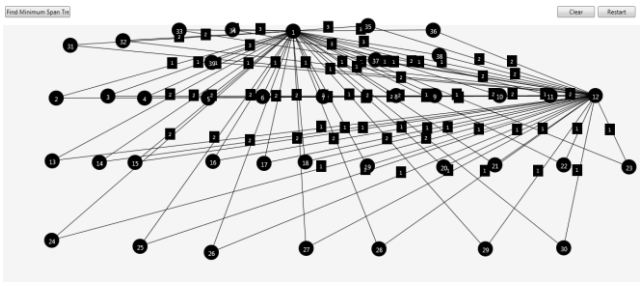

Fig. 9 Network Graph of Wiki Data based on Geodesic Distance

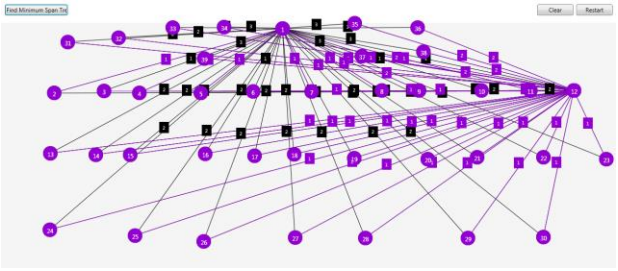

Fig. 10 Minimum Spanning Tree of Wiki Network

5. Conclusion

Social learning network has illustrated the model of interaction between student and system as social-relationship with different level of interaction intensity. The relationship was modeled as matrix adjacency with matrix for geodesic distance. Based on these two matrices, hierarchical clustering and graph clustering (minimum spanning tree) have been utilized to analyze the behavior of student. According to result from experimental result we have succeeded of proving our hypothesis between outdegree and indegree value with cluster characteristic. High outdegree student will drag another student with high degree value to be at the same cluster, this also applied to student who has high indegree value. In the MST network the shortest path will represent active student that has fast network to broadcast the information within network. It also can be called as a broker or relay that can supply any information within the network. This finding will bring great benefit on e-learning system and its users. Due to most of previous researcher still focus on manual selection of each actor.

Acknowledgments

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] UCINET. 2011. UCINET Documentation [Online]. [Accessed 2011].

[2] Tan, P.-N., Steinbach, M. & Kumar, V. 2006. Introduction to Data Mining, Boston, MA, Pearson International Edition.

[3] Yu, G. 2007. Social Network Analysis Based on BSP CLustering Algorithm. Communications of the IIMA, Volume 7, Issue 4.

[4] Handcock, M. S., Raftery, A. E. & Tantrum, J. M. 2007. Model-Based Clustering for Social Networks. J. R. Statist. Soc. A (2007)170, Part 2, pp. 301–354.

[5] Mishra, N., Schreiber, R., Stanton, I. & Tarjan, R. E. Year. Clustering Social Networks. In: WAW, 2007. Springer.

[6] Agarwal, N., Galan, M., Liu, H. & Subramanya, S. 2010. WisColl: Collective wisdom based blog clustering. Information Sciences, Elsevier 180 39-61.

[7] Grissa,D., Guillaume,S. & Nguifo,E.M. 2010. Combining Clustering techniques and Formal Concept Analysis to characterize Interestingness Measures., Complexe scientifique des.Cézeaux, , France. Last Accessed 2015. [Accessed].

[8] Guojun Gan and Jianhong Wu. 2004. Subspace clustering for high dimensional categorical data. SIGKDD Explor. Newsl. 6, 2 (December 2004), 87-94. DOI=http://dx.doi.org/10.1145/1046456.1046468 .

[9] Hattie, John. (2002). Schools Like Mine: Cluster Analysis of New Zealand Schools.

[10] Hsu, Jia-Lien & Yang, Hong-Xiang. 2009. A Modified K-means Algorithm for Sequence Clustering. Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009. 1. 287-292. 10.1109/HIS.2009.64.

[11] Mark Junjie Li, Michael K. Ng, Yiu-ming Cheung, and Joshua Zhexue Huang. 2008. Agglomerative Fuzzy K-Means Clustering Algorithm with Selection of Number of Clusters. IEEE Trans. on Knowl. and Data Eng. 20, 11 (November 2008), 1519-1534. DOI: https://doi.org/10.1109/TKDE.2008.88.

[12] Park, Yeonjeong & Hyun Yu, Ji & Jo, Il-Hyun. (2015). Clustering blended learning courses by online behavior data case study in a Korean higher education institute. The Internet and Higher Education. 29. 10.1016/j.iheduc.2015.11.001.

[13] R. Halverson, Lisa & Graham, Charles & J. Spring, Kristian & S. Drysdale, Jeffery & R. Henrie, Curtis. (2014). A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. Internet and Higher Education. 20. 20-34. 10.1016/j.iheduc.2013.09.004..

[14] Shiau, W.-L., Y. K. Dwivedi, et al. (2017). "Co-citation and cluster analyses of extant literature on social networks." 37(5): pp.390-399.

[15] Chai, Sangmi & Kim, Minkyun. (2012). A socio-technical approach to knowledge contribution behavior: An empirical investigation of social networking sites users. International Journal of Information Management - INT J INFORM MANAGE. 32. 10.1016/j.ijinfomgt.2011.07.004.

[16] K.W. Chu, Samuel & Capio, Catherine & Aalst, Jan & Cheng, Eddie W.L.. (2017). Evaluating the use of a social media tool for collaborative group writing of secondary school students in Hong Kong. Computers & Education. 110. 10.1016/j.compedu.2017.03.006.

[17] Mansur,A.B.F., Yusof,N., Basori,A.H., 2017, Comprehensive analysis of Student's Academic Failure Classification through Role-Sphere Influence and Flow betwenness centrality, Procedia Computer Science,Vol.116,pp.509-515.

Andi Besse Firdausiah Mansur, received B.Sc(Mathematics), in 2004 and MSc(Software Engineering) in 2009. Furthermore, she obtained Ph.D. degree in Software Engineering from Universiti Teknologi Malaysia, Johor Bahru, Johor, in 2013. From 2013 to present, she is assistant professor at Faculty of Computing and Information Technology Rabigh, King Abdulaziz University. She is member of reviewer board in some international journal. Her research interest: Social Network Analysis, E-learning, Clustering, Data Mining and Mathematics

Norazah Yusof is an Associate Professor at the Faculty of Computing and Information Technology Rabigh, King Abdulaziz University. She received her B. Sc.(Computer Science) and MSc (Computer Information Systems) from University Of Miami, Florida, USA and her PhD (Systems and Science Management) from Universiti Kebangsaan Malaysia in 2005.Her research interests include Social Network Analysis, Learning Analytics, Big Data Analytics, Soft Computing, Teamwork effectiveness, Ontologies and Software Engineering..

Nada Omar Bajnaid. B. Sc of Computer Science. MSc (Computer Science), University of Wisconsin Milwaukee. Ph. D (Ontology and Semantic Web), 2014, London Metropolitan University. Nada is assistant professor at King Abdulaziz University. Her research interests are Ontology, Semantic Web, Context Awareness and software quality.