Development of Self-Organizing Map Clustering Algorithm for Relay Selection in High Density Long Term Evolution Networks

Nwaogu Onyeyirichi E.[1]
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria

Ozuomba Simeon[2]
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria
simeonoz@yahoomail.com
simeonozuomba@uniuyo.edu.ng

Kalu Constance[3]
Department of Electrical/Electronic and Computer Engineering, University of Uyo, Akwa Ibom, Nigeria

***Abstract*— In this paper, Long Term Evolution (LTE) network with high density of nodes is studied. In such networks, relays are required to improve on the efficiency of signal transmission among the nodes, enhance the Quality of Service (QoS) and manage the communications emanating from the teeming LTE node population. In the study, six hundred (600) Long Term Evolution (LTE) network nodes that are randomly located in an area of about 600 m by 600 m with the base station located at the center of the region were considered. The location of data and the energy dissipation of data from the 600 nodes were generated using a random number generation tool in MATLAB. Specifically, forty (40) LTE nodes located between 140 m to 160 m meters were considered as prospective relays. Self Organizing Map (SOM) was utilized in the selection of the prospective relays to determine the actual relays (cluster heads) for the process and nine (9) LTE nodes from the prospective relays were identified as cluster heads with the other nodes being considered as cluster slaves. The Self Organizing Map (SOM) clustering algorithm was used to cluster the cluster slaves to the cluster heads. From the results, the most effective cluster head had seventy nine (79) cluster slaves with the least having 39 clusters. The pattern for relay selection with SOM and a literature reviewed algorithm LEACH was compared and LEACH selected only three (3) cluster heads from the LTE nodes as against the SOM clustering algorithm that selected nine (9) cluster heads from the LTE nodes. As such, SOM clustering was the better in relay selection. It is therefore recommended that LTE wireless network service providers should adopt the SOM algorithm developed in this research for real time relay selection especially in the highly dense region of their networks.**

Keywords*— Long Term Evolution (LTE) networks, Clustering, Relay selection, Relay Heads, Cluster slaves, Base station, Self-Organizing Map (SOM), Low-Energy Adaptive Clustering Hierarchy (LEACH).

I. INTRODUCTION

Long Term Evolution (LTE) networks have been widely deployed in cellular communications to provide high-speed data for mobile phones and smart devices [1,2]. With the increasing number of mobile users and demand for high data rates, density in LTE networks becomes a major issue especially in crowded areas such as large churches, conferences, and concerts [3,4]. Proper frequency allocation may not be sufficient for mobile users due to the shortage of the spectrum [5,6,7]. This congestion in a limited area cause the Base station (BS) to be overloaded and thus not able to serve all users successfully. Therefore, the main objective of this paper is to develop clustering algorithm for relay selection in high density LTE networks without the need of additional infrastructure. In this case, some of the LTE nodes will be selected to act as relay heads (ad-hoc base station) [8,9,10]. The selected relay heads will then relay all other LTE nodes that are in the same cluster to the base station [11,12,13]. These cluster members are called cluster slaves [14,15].

Clustering is a form of machine learning method that entails grouping of data points into classes of groups such that the data points in each group are more similar to one another [16,17]. One of the simplest clustering algorithm is the K-means which is an unsupervised machine learning algorithms that can be used to group data points in a dataset into different clusters (groups) [16,17,18,19]. However, studies have identified some flaws in the k-means approach [20,21]. Among other issues, in K-means the number of expected cluster must be specified in advance, it is not deterministic as the solution depends on the initialization value [20,21]. Hierarchical clustering techniques have been also used to address some of the issues in K-means method [20]. Particularly, Hierarchical Agglomerative Clustering has been applied in Wireless Sensor Networks (WSNs) and in many other disciplines. Ma, Fang and Bai [22] studied relay-based clustering algorithm (RBC), Low-Energy Adaptive Clustering Hierarchy (LEACH), LEACH energy efficient (LEACH-E) and SEP-M algorithms. The results of their study showed that LEACH has limited effect on balancing energy consumption whereas LEACH-E, SEP-M and RBC perform better than LEACH over time, since the three algorithms achieved the effect of balancing energy consumption. However, RBC outperforms all the other algorithms.

In this study, relay nodes were selected with Self-Organizing Map (SOM) clustering technique which is a form of artificial neural network method. The relay nodes selection process of SOM was compared with that of Low-Energy Adaptive Clustering Hierarchy (LEACH), which is another clustering technique already used in similar relay selection study. The effectiveness of the relay node (relay) selection by the clustering techniques was compared in terms of number of relays that were selected from a given network nodes. The SOM model was further used to cluster the LTE nodes to the relay heads to determine the number of cluster slaves (cluster slaves are the other LTE nodes not considered as relay nodes) contained by each of the relay heads (relays).

II. METHOD

In this study, relay nodes were selected with Self-Organizing Map (SOM) procedure. Six hundred Long Term Evolution (LTE) nodes were assumed to occupy a region of 600 square meters. The base station (denoted as sink) was assumed to be at the centre of the entire network to ensure maximum signal transmission of all the nodes to the sink. The selection process of SOM was compared with that of LEACH algorithm. The effectiveness of the relay node (Relay) selection by the proposed models was compared in terms of number of relays that were selected. The SOM model was further used to cluster the LTE nodes to the relay (cluster heads) to determine the number of cluster slaves (cluster slaves are the other LTE nodes not considered as relay nodes) contained by each of the relay nodes. The flow diagram for the research process is displayed in Figure 1.

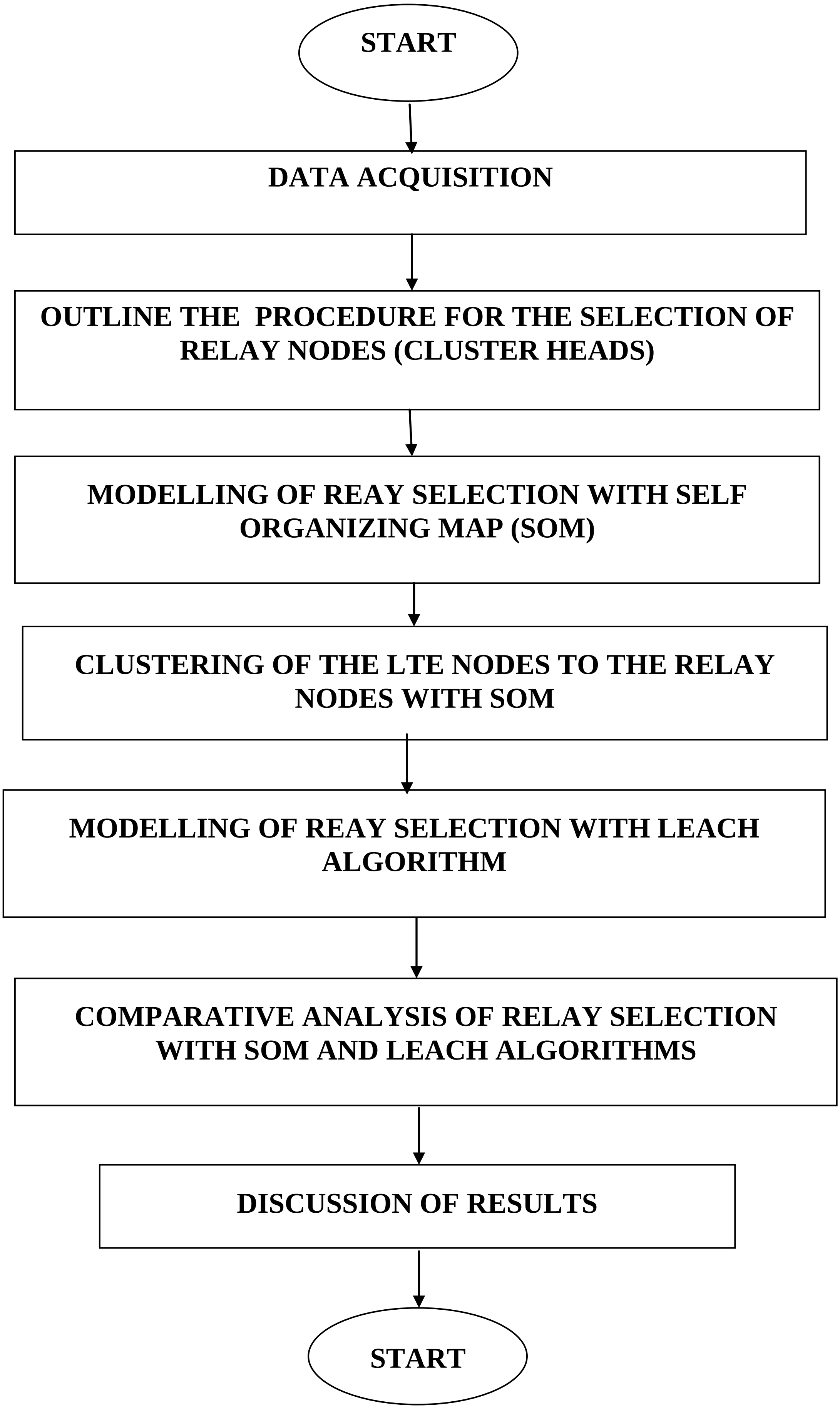

Figure 1; Flow diagram of the Research Procedure

A. Acquisition of the study data

In the study, six hundred (600) LTE network nodes that are randomly located in an area of about 600 m by 600 m with the base station located at the centre of the region, as shown in Figure 1.

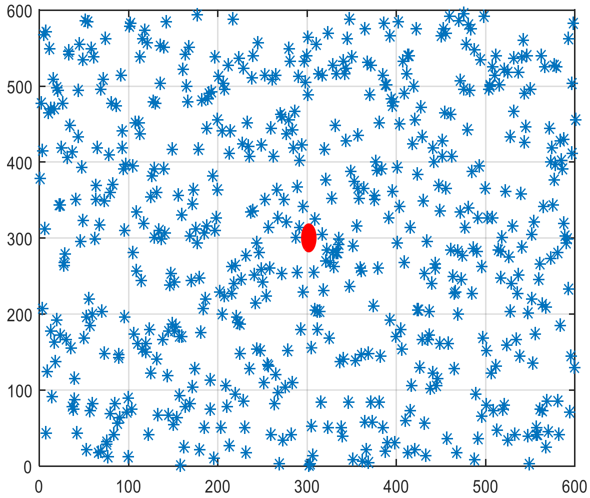

Figure 1; Position of the LTE nodes.

The location data and the energy dissipation data of the 600 nodes are generated using a random number tool in MATLAB. The random numbers for the location (x and y coordinates) of each of the 600 LTE nodes are generated based on normal distribution probability function given by the expression:

$$randnd = lb + (ub - lb) \times rand(600,1) \quad (1)$$

Where $randnd$ is the randomly generated data for normal distribution, lb is the lower boundary which indicates the minimum point of the location of the devices from the base station (0 mx^2

$$\mu = \frac{ub+lb}{2} \quad (2)$$

The location (x and y coordinates) generated were used to compute the position or the distance (d) of each of the LTE nodes from the origin of the region, as shown in Figure 1. The Red point in Figure 1 is the location of the base station; it is at point 300 x 300 (x and y axis of 300). This is necessary to ensure a centralized transmission of signal to all the LTE nodes.

The distance from the base station to the location of a node is computed using the coordinate geometry formula for the distance between two points (x1, y1) , (x2, y2) and it is given as;

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (3)$$

According to research findings, several Long Term Evolution (LTE) nodes cannot transmit efficiently due to their distance from the sink (base station) or their rate of energy dissipation [23]. As such an intelligent clustering algorithm was needed to properly select the LTE nodes that are closer to the sink with appropriate energy dissipation balance that can act as relay nodes (relay) for faster and efficient signal transmission from the sink to the other LTE nodes. The rate of energy dissipated is dependent on the distance of the node from the base station along with other peculiar features of the device and it is computed as follows;

$$E_{AT} = (Ee + E_{DA})k + E_{amp}k\sqrt{d_{bs}} \quad (4)$$

Where E_{AT} is the rate of energy dissipated after transmission, Ee is the transmitted energy per bit, E_{DA} is the data aggregation energy, E_{amp} is the transmit amplifier energy, k is the size of package per LTE node and d_{bs} is the distance of the LTE node from the base station. The typical data used for computing the rate of energy dissipated after transmission are shown in Table 1.

Table 1. The data used for computing the rate of energy dissipated after transmission

Model Parameters	Values
K	Standard data in bits is 4000bits
E_e	Standard value of a node is 2Joules
E_{DA}	5×10^{-9}J
E_{amp}	100×10^{-12}J

The data for the values of Ee (energy transmitted per bits) for each of the node is generated using the ''rand'' command in MATLAB 2015a. The random numbers for Ee of each of the 600 LTE nodes is generated based on normal distribution probability function given by the following expressions:

$$randnd = l\text{Ee} + (u\text{Ee} - l\text{Ee}) \times rand(600,1) \quad (5)$$

Where $randnd$ is the randomly generated data for normal distribution, $l\text{Ee}$ is the lower boundary which indicates the minimum value of E_e of the devices and uE_e is the upper value which indicates the maximum value of E_e of the device. The mean (μE_e) of the distribution is given as;

$$\mu\text{Ee} = \frac{u\text{Ee} + l\text{Ee}}{2} \quad (6)$$

A detailed relay selection was done to properly select the relay nodes that can properly transmit signal (act as intermediary) between the base station and the cluster slaves (which are the other LTE nodes not considered as relay nodes). The randomly generated values of Ee are used along with the specified values of E_{DA} , E_{amp} and K to compute E_{AT} which is the rate of energy dissipated by each of the nodes.

B. Relay Selection with Self-Organizing Map (SOM) and Leach clustering algorithms

Specifically, Self organizing map (SOM) and Leach clustering algorithms are considered in this study. SOM is the clustering algorithm of the artificial intelligence which is essential in clustering and further classification of the data to their relays heads. The self-organizing map (SOM) is a sub-toolbox under the artificial intelligent toolbox in MATLAB. Leach algorithm is from existing work and it is implemented here so as to compare the results of the SOM with that of an existing mechanism. The detailed step involved in the selection and clustering of the LTE nodes using the SOM and Leach algorithms is given in Figure 2.

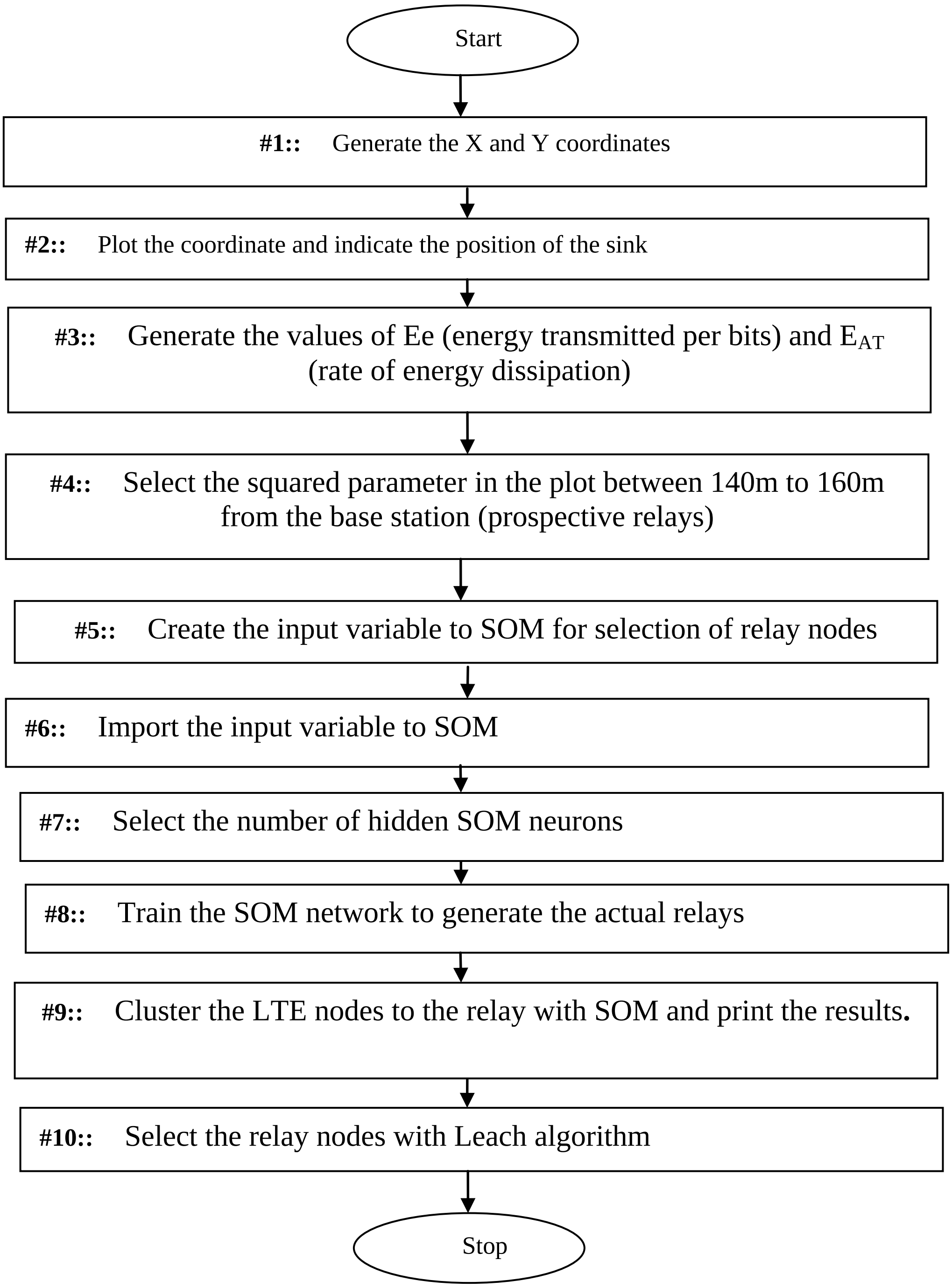

Figure 2 The detailed step involved in the selection and clustering of the LTE nodes using the SOM and Leach algorithms

Step One and two: Generate the x and y coordinates of the LTE nodes and plot the location of the nodes

In step one, the location of each of the 600 LTE nodes is generated using the random number expression given in Equation 1 and also to plot the location of the nodes based on the Matlab code in Figure 3. In the case of Figure 3, the value of $ub = 600$ and $lb = 0$. Then, the mean of the distribution according to Equation 2 it is $\mu = \frac{600+0}{2} = 300$.

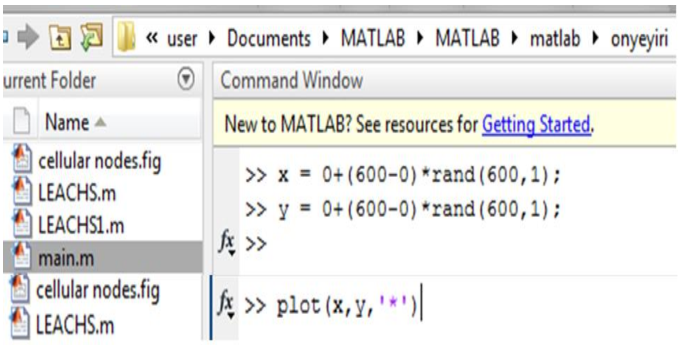

Figure 3: Generation of the X and Y coordinates

Step three:: Generate the values of Ee (energy transmitted per bits) and E_{AT} (rate of energy dissipation)

The EAT (rate of energy dissipation) is determined based on Equation 4. The value of the Ee (energy transmitted per bits) is randomly generated with normal distribution based on Equation 5 whereas the values of the other parameters for computing the rate of energy dissipation are obtained from Table 1. The code in Matlab for generating the Ee and computing the EAT (rate of energy dissipation) is as shown in Figure 4.

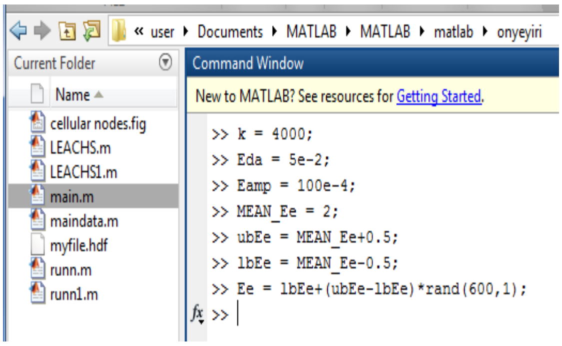

Figure 4: Generate the values of Ee (energy transmitted per bits) and E_{AT} (rate of energy dissipation)

Step Four:: Select the squared region within the plot between 140 m to 160 m from the base station (prospective relays)

The region of interest for the prospective relays (cluster heads) is first selected. In this research, the selected region lies between 140 m to 160 m from the base station as shown in Figure 5.

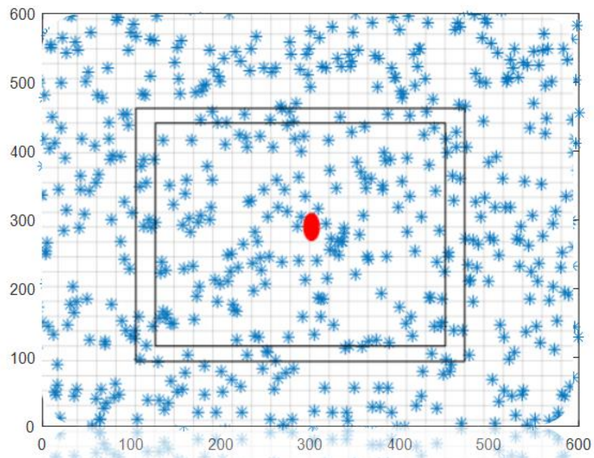

Figure 5; Selection of possible relay LTE nodes (cluster heads) within a region of 140 to 160 square meter from the base station.

Step Five:: Create the input variable to the SOM algorithm for selection of the relay nodes

Before the input data can be used for the SOM algorithm, a data variable (in this case, r as shown in Figure 6) need to be created. The variable r will contain the two column matrix with one column containing the generated value of d and the other column containing the generated value of Eat, as shown in Figure 6. At this point, the GUI of the SOM toolbox in MATLAB is called up using the Matlab nctool command, as shown in Figure 6. The GUI of the SOM toolbox in MATLAB ; the architectural network of the SOM has the input layer, the hidden SOM neuron layer and the output layer.

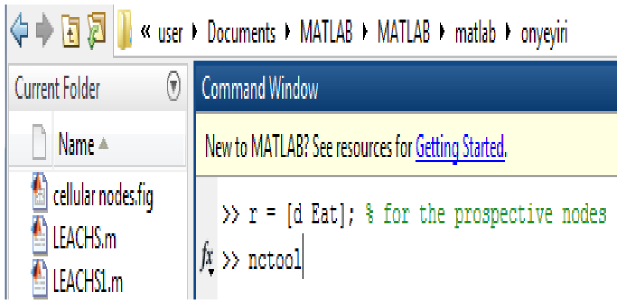

Figure: 6 Input variable to SOM for selection of relay nodes

Step Six and Seven: Import the input variable to SOM work area in Matlab and select the number of hidden SOM neurons

The input variable (r) created in step five already contains the 600 rows two column matrix with one column containing the generated value of d and the other column containing the generated value of Eat, as shown in Figure 6. At this point the 600 x 2 matrix dataset store as variable r is loaded as input for the SOM and Leach clustering algorithm operations.

In this study, six (6) hidden neurons are selected along with two input neurons and 36 output neurons. It need to be noted that when six hidden neurons are selected for SOM, it is used as dimension value which the algorithm use to generate a square of the dimension size, as shown in Figure 7.

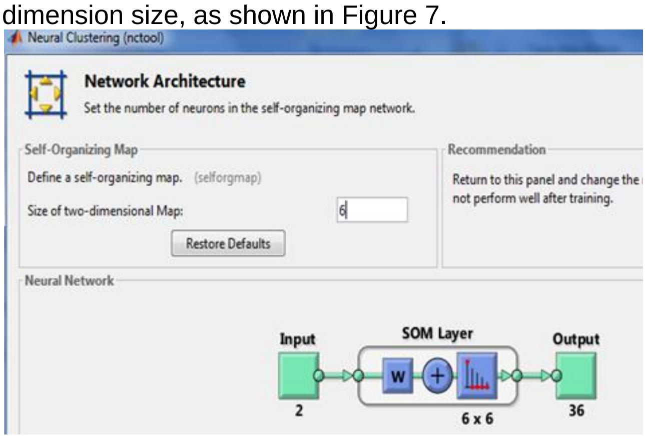

Figure 7: The selection of the number of neurons for the three sections of the SOM architecture

Step Eight and Nine: Train the SOM network to generate the actual relays and Cluster the LTE nodes to the relay using SOM algorithm

After the data is loaded and the neurons are set, the SOM algorithms is trained and retrained until acceptable minimal error value is obtained. At this point, the relays nodes or cluster heads are identified from within the specified region which lies between 140 m to 160 m from the base station as shown in Figure 5. Next, the other nodes around the base station are then clustered to the pre-selected relays (cluster heads). In this case, for the SOM algorithm three (3) hidden neurons are selected along with two input neurons and 9 output neurons, as show in Figure 8. After the data is loaded and the neurons are set, the SOM algorithms is trained and retrained until acceptable minimal error value is obtained. At the end of the training, the nodes are clustered to the earlier selected relays.

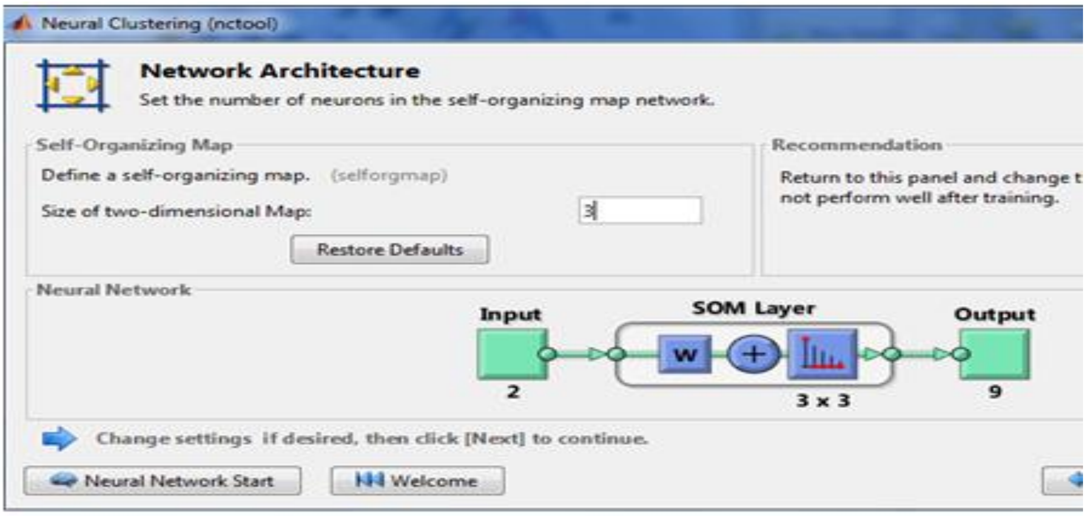

Figure 8; The selection of the number of neurons for the three sections of the SOM architecture based on the number of selected relays in the step eight (8)

Step Ten:: Select the relay nodes with the LEACH algorithm

Since the data is already loaded in the Matlab, the relevant Matlab code for the LEACH clustering algorithm are used to implement the selection of the relay and at the same time clustering of the nodes to the relays. The screenshot for the Matlab code used to implement the LEACH clustering algorithm is shown in Figure 9.

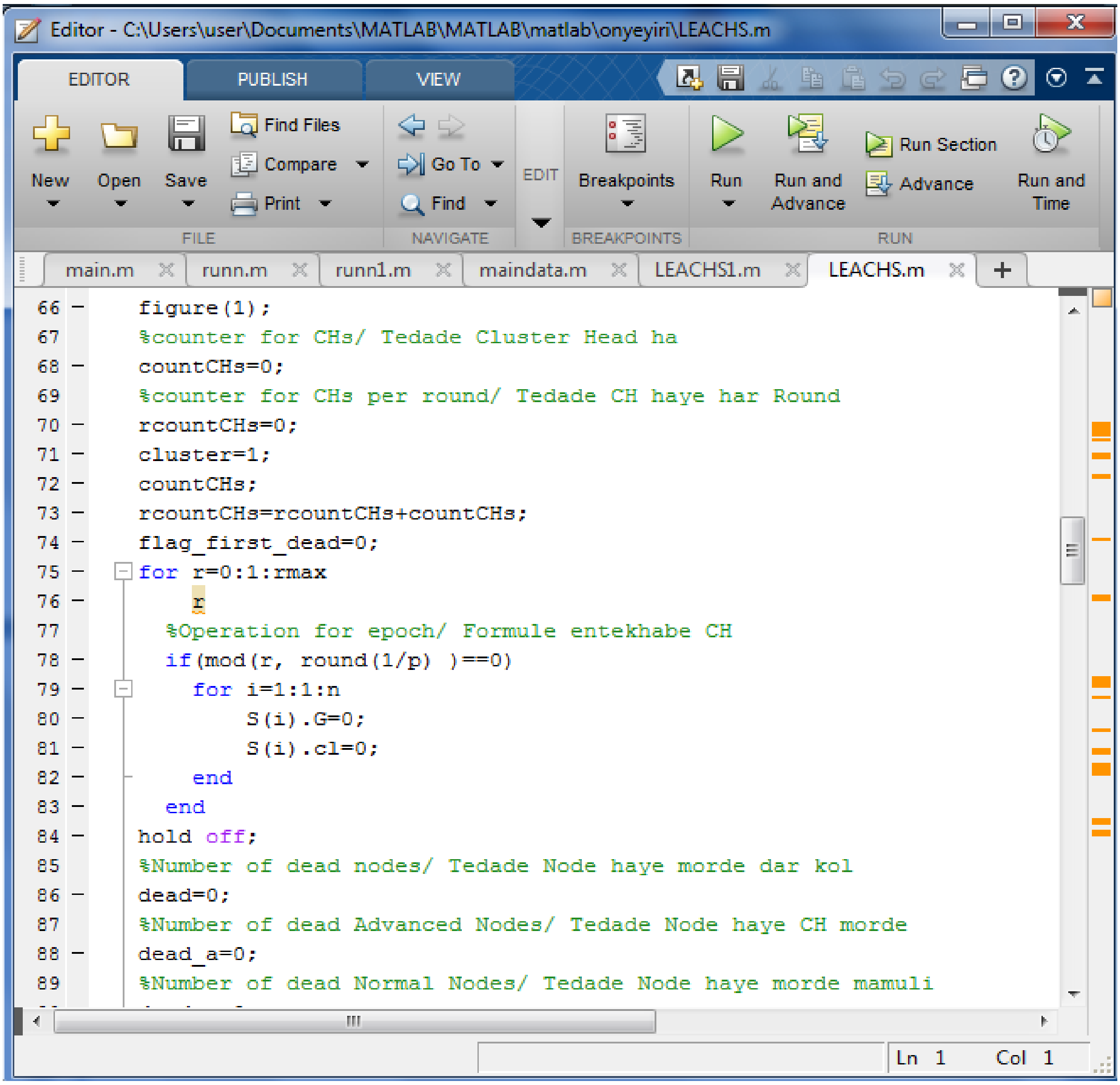

Figure 9: Selection of relay nodes and clustering of the nodes to the relays using LEACH Algorithm

III. RESULTS

The nodes that are within the specified region of 140 m to 160 m from the base station are shown in Table 2 along with their x and y coordinates and their energy dissipation data. In all, about 40 nodes were found within the specified region and those are the prospective cluster heads.

Table 2: The selected prospective cluster heads with their the x and y coordinates and energy dissipation

S/N	LTE nodes	X-coordinate (m)	Y-coordinate (m)	Rate of energy dissipation x 10^{-6}J
1	34	448.32	201.32	6574.1
2	35	156.31	88.373	6711
3	72	147.64	596.45	7451
4	88	142.72	256.32	9426.9
5	89	146.19	589.1	7921.6
6	93	440.25	495.45	8499.5
7	107	452.72	46.841	7491.7
8	116	153.26	184.21	9735.5
9	120	150.05	49.196	8103.6
10	131	140.02	36.008	8114.9
11	187	443.69	270	7717.7
12	200	456.26	279.4	9477.2
13	204	440.32	69.594	8435.8
14	214	440.26	223.84	7278.9
15	224	148.21	567.03	7064.1

16	233	451.92	133.95	8756.1
17	244	153.58	422.76	6556
18	247	152.4	153.92	8365.3
19	259	444.78	569.89	7717.4
20	271	458.87	346.89	6655.8
21	278	151.84	15.012	7675.1
22	280	454.92	556.44	7099.5
23	284	442.24	535.94	8307.9
24	305	156.44	204.29	8449.8
25	331	447.71	131.41	7513.6
26	360	146.5	11.232	7043.1
27	368	148.16	451.87	10249
28	381	451.16	471.75	7411.8
29	382	145.07	524.09	7580.8
30	403	449.09	204.59	8716.5
31	406	153.67	220.69	9192.6
32	421	454.7	448.93	9516.9
33	479	443.41	45.036	7970.6
34	500	152.25	187.57	8847.4
35	523	159.18	364.82	7838.3
36	538	143.93	477.29	9046.2
37	545	445.93	312.65	6482.6
38	571	444.32	27.085	9534.8
39	595	144.25	360.36	9015.7
40	600	456.21	72.286	6659.9

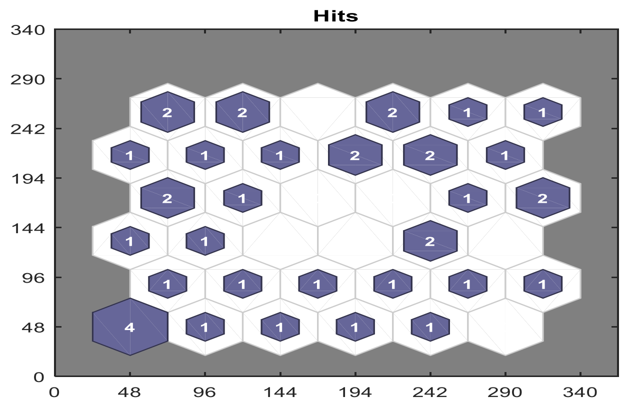

Figure 10: clustering for the cluster heads

From the 40 the prospective cluster heads 9 nodes were selected as cluster heads by SOM algorithm whereas 3 nodes were selected as cluster heads by the LEACH algorithm. The pattern for the selection of cluster was 100% because the total number (40) of the prospective cluster heads was supplied as input to the SOM model and only 9 were selected as actual cluster heads. Furthermore, the SOM was used to cluster the rest of the odes to the 40 selected cluster heads and the result is shown in Figure 11.

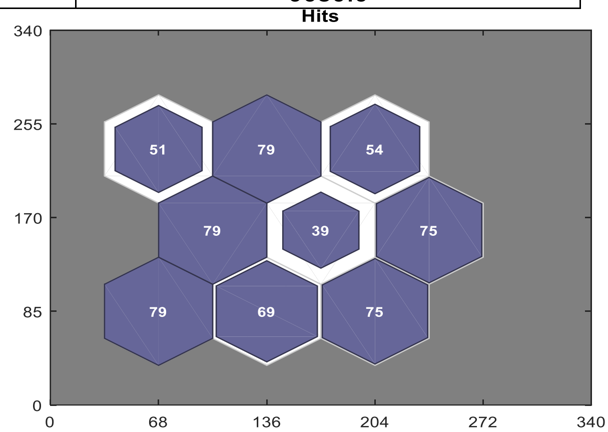

Figure 11: Number of cluster slaves in each cluster heads

From Figure 11 it be seen that three (3) relays had the highest number of clusters of 79 each (namely, relays 1, 4 and 8) as also shown in Figure 12. Also, the cluster head with the least cluster slaves was the 5th relay with 39 cluster slaves, as also shown in Figure 12.

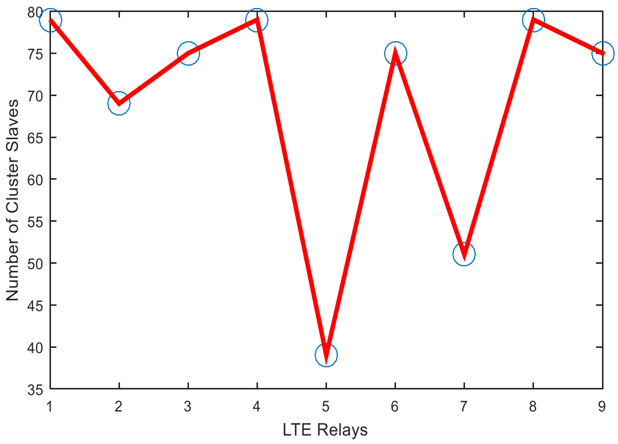

Figure 12: Clustering of LTE nodes per Relay.

Fang (2018), proposed a Low-Energy Adaptive Clustering Hierarchy (LEACH), for the selection of relays and the result after simulation is shown in Figure 4.11. According to Fang(2018) [23] any node that transmits up to 1000 times was considered as a relay node and from Figure 13, only 3 nodes achieved the required condition. If the remaining 597 cluster slaves were assigned to the 3 cluster heads the efficiency of the transmission will drop drastically. As such, the higher the number of cluster heads, the better the efficiency of the transmission network. The bar chart in Figure 14, summarizes the number of cluster heads selected by each of the two cluster algorithms (SOM and LEACH) considered in this paper.

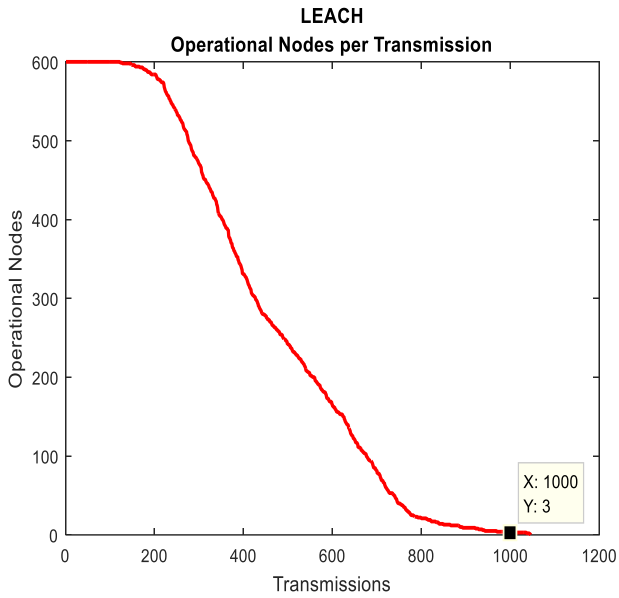

Figure 13; Relay Selection with LEACH

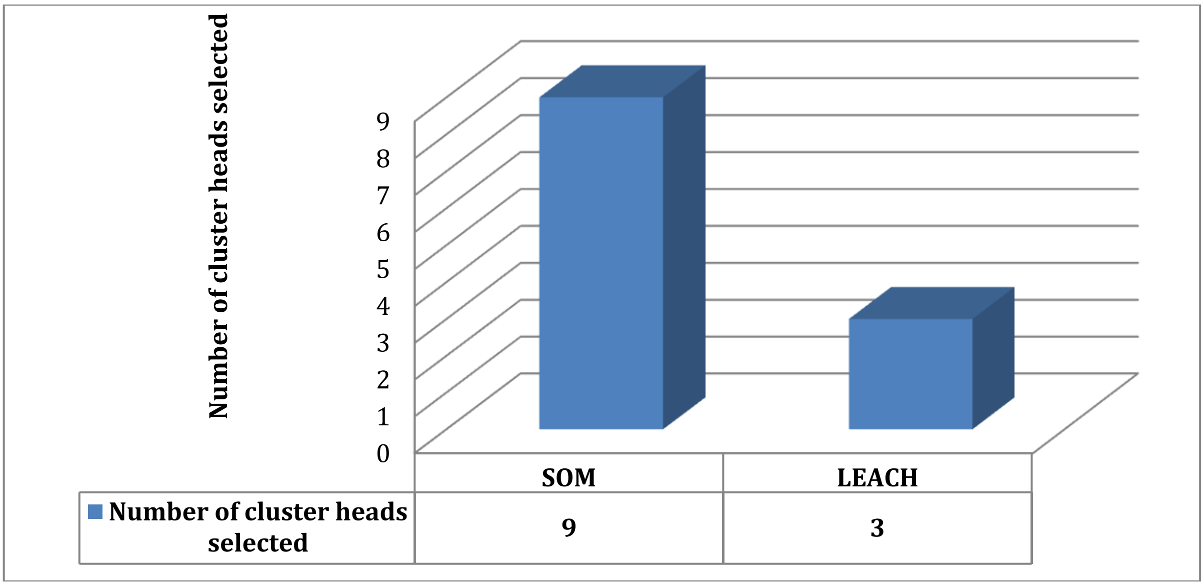

Figure 14; Comparison of number of cluster heads selected by SOM and LEACH cluster algorithms.

IV. CONCLUSION

Due to the constant increase of LTE node users, it is imperative to select relays from the existing nodes for ease of transmission with the base stations. As such, in this paper, SOM cluster algorithm was used to carry out the selection of cluster heads for a case study LTE network. The relay selection capability of SOM is compared with that of LEACH algorithm to determine the best model for relay selection. SOM was able to select 9 relays as against LEACH that had 3 relays. As such, SOM clustering was the better in relay selection. Also, SOM clustering model was used in carrying out clustering of the other LTE nodes to the selected relays (cluster heads). It is recommended that LTE wireless network service providers should adopt the SOM algorithm developed in this research for real time relay selection especially in the highly dense region of their networks.

REFERENCES

1. Gozalvez, Javier. "First Commercial LTE Network [Mobile Radio]." *IEEE Vehicular Technology Magazine* 5.2 (2010): 8-16.
2. Dang, Shuping, et al. "From a Human-Centric Perspective: What Might 6G Be?." *arXiv preprint arXiv:1906.00741* (2019).
3. Eschbach, Benedikt. *Options for the Joint Evolution of Digital Terrestrial Television and Mobile Internet Access*. Wissenschaftsverlag Mainz, 2015.
4. Davidson, Charles M., and Michael J. Santorelli. "Understanding the debate over government-owned broadband networks." *Advanced Communications Law & Policy Institute, New York Law School, June* (2014).

5. Matheson, Robert, and Adele C. Morris. "The technical basis for spectrum rights: Policies to enhance market efficiency." *Telecommunications Policy* 36.9 (2012): 783-792.
6. Wellenius, Bjorn, and Isabel Neto. *Managing the radio spectrum: framework for reform in developing countries*. The World Bank, 2008.
7. Ravi, Shamika, and Darrell M. West. "Spectrum policy in India." *Center for Technology Innovation at Brookings* (2015).
8. Uyoata, Uyoata, and Mqhele Dlodlo. "Relay Assisted Device-to-Device Communication: Approaches and Issues." *arXiv preprint arXiv:1810.07799* (2018).
9. Hajjar, Maryam, et al. "Relay selection based clustering techniques for high density LTE networks." *Wireless Networks* 25.5 (2019): 2305-2314.
10. Hajjar, Maryam, et al. "Hybrid clustering scheme for relaying in multi-cell LTE high user density networks." *IEEE Access* 5 (2017): 4431-4438.
11. Zhao, Yifeng, et al. "Distributed Dynamic Cluster-Head Selection and Clustering for Massive IoT Access in 5G Networks." *Applied Sciences* 9.1 (2019): 132.
12. Huang, Hongcheng, et al. "User-information-aware D2D multicast file distribution mechanism." *Sensors* 18.10 (2018): 3389.
13. Huang, Jun, et al. "Network coding in relay-based device-to-device communications." *IEEE network* 31.4 (2017): 102-107.
14. Wulandari, Diah Ayu Retnani, and Mochamad Edoward Ramadhan. "Performance comparison analysis library communication cluster system using merge sort." *Journal of Physics: Conference Series*. Vol. 1008. No. 1. IOP Publishing, 2018.
15. Brust, Matthias R., Adrian Andronache, and Steffen Rothkugel. "Waca: A hierarchical weighted clustering algorithm optimized for mobile hybrid networks." *2007 Third International Conference on Wireless and Mobile Communications (ICWMC'07)*. IEEE, 2007.
16. Amruthnath, Nagdev, and Tarun Gupta. "Fault Diagnosis using Clustering. What Statistical Test to use for Hypothesis Testing?." *arXiv preprint arXiv:1904.13365* (2019).
17. Berkhin, Pavel. "A survey of clustering data mining techniques." *Grouping multidimensional data*. Springer, Berlin, Heidelberg, 2006. 25-71.
18. Forsberg, Fredrik, and Pierre Alvarez Gonzalez. "Unsupervised Machine Learning: An Investigation of Clustering Algorithms on a Small Dataset." (2018).
19. Ullman, J. D., and A. Rajaraman. "Clustering." *Mining of Massive Datasets* (2012): 241-280.
20. Adams, Ryan P. "Hierarchical Agglomerative Clustering." *Proc. Ninth SIAM Data Mining Conf.(SDM'09)*. 2009.
21. Su, Ting, and Jennifer Dy. "A deterministic method for initializing k-means clustering." *16th IEEE International Conference on Tools with Artificial Intelligence*. IEEE, 2004.
22. Ma, Xiaofu, Yu Fang, and Xingzhen Bai. "A balanced energy consumption clustering algorithm for heterogeneous energy wireless sensor networks." *2010 IEEE International Conference on Wireless Communications, Networking and Information Security*. IEEE, 2010.
23. Fang, Y., Ma, X., and Jiang, M. (2018): a relay-based clustering algorithm for heterogeneous energy wireless sensor networks. Shangai, China.