INVESTIGATION OF ENERGY DEMAND DATA CORRELATION TO SIGNIFICANT VARIABLES DURING MOVEMENT CONTROL ORDER

SHARIZAD BIN SAHARANI

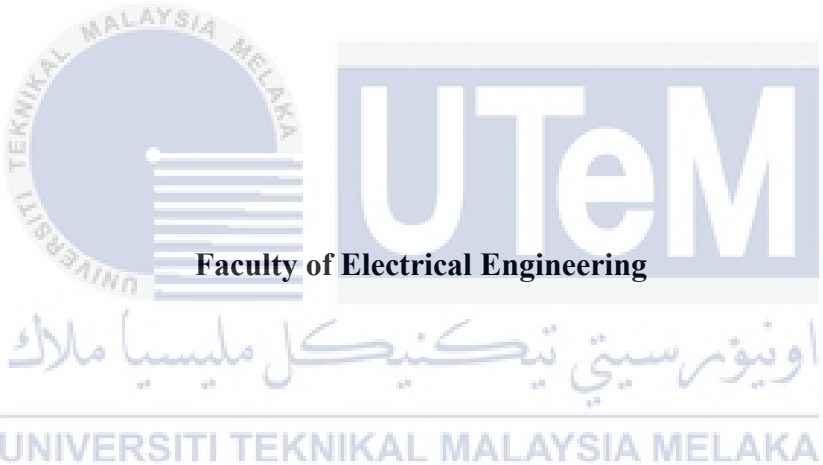
BACHELOR OF ELECTRICAL ENGINEERING WITH HONORS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

Investigation of Energy Demand Data Correlation to Significant Variables During Movement Control Order

SHARIZAD BIN SAHARANI

A report submitted
in partial fulfillment of the requirements for the degree of
Electrical Engineering



Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I declare that this thesis entitled “Investigation of Energy Demand Data Correlation to Significant Variables During Movement Control Order” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : 

Name : SHARIZAD BIN SAHARANI

Date : 5/7/2021

APPROVAL

I hereby declare that I have checked this report entitled “Investigation of Energy Demand Data Correlation to Significant Variables During Movement Control Order” and in my opinion, this thesis it complies the partial fulfillment for awarding the award of the degree of Bachelor of Electrical Engineering with Honours

Signature : .....

Supervisor Name : Ts. Dr. Mohamad Fani bin Sulaima

Date : 5/7/2021

DEDICATION

I would like to dedicate my thesis to my beloved family without whose constant support this thesis paper was not possible

ACKNOWLEDGEMENT

Assalamualaikum w.b.t praise upon Allah for His gifts and blessing, which help us complete this study successfully. First of all, I would like to express our gratitude to our supervisor, Ts. Dr. Mohamad Fani Bin Sulaima for his assists and guide from the start until the end of this study completion.

Also, the highest gratitude to our parents and families for supporting us morally and financially. Without their aid, I will never complete this study, and I would have given up on this project.

Besides, I also wish a million thanks to all my friends, especially my classmates, for sharing their knowledge, skills, time, and equipment with us. Lastly, my million thanks to everyone associated with this study, either directly or not. These people contributions gave me the will and determination to finish this project and complete this report.

ABSTRACT

The worlds currently face a significant impact from the covid-19 pandemic, which also influences energy consumption. This study investigates the substantial connection of the classified data between energy consumption, cooling degree days, average temperature, and covid-19 cases information using mathematical and neural network approaches, which are regression analysis and self-organizing maps. It is well established that various data mining methods have revamped the classification process of data analytics. Specifically, this study investigates the correlation between the collected variables using regression analysis and selecting the best-matching unit under the normalization method using self-organizing maps. The self-organizing maps become better when the datasets have variations; the result denotes that this method produced high mapping quality based on the map size and normalization method. The data crossing connection is indicated by using the regression analysis method and the classified data results during the movement control order are validated in self-organizing maps to achieve the study objective. By performing these methods, this study established that the correlation between the energy demand towards cooling degree days, average temperature, and covid-19 cases is very weak and the verification of *'logistic'* and *'range'* normalization method produced the best classification result.

Abstract should consists of a brief on the method use, findings & conclusion.

ABSTRAK

Dunia saat ini menghadapi kesan ketara disebabkan pandemik covid-19, yang juga mempengaruhi penggunaan tenaga. Kajian ini bertujuan untuk mengkaji hubungan kepada data terklasifikasi antara penggunaan tenaga, hari darjah pendinginan, suhu purata dan maklumat kes covid-19 menggunakan pendekatan rangkaian matematik dan rangkaian neural, yang merupakan analisis regresi dan peta pengorganisasian diri. Sudah terbukti bahawa pelbagai kaedah perlombongan data telah mengubah proses pengkelasan analisis data. Secara khusus, kajian ini menyelidiki perkaitan antara pemboleh ubah yang dikumpulkan menggunakan analisis regresi dan pemilihan unit pemadanan terbaik di bawah kaedah normalisasi menggunakan peta pengorganisasian diri. Peta pengorganisasian diri menjadi lebih baik apabila kumpulan data mempunyai variasi, hasilnya menunjukkan bahawa kaedah ini menghasilkan kualiti pemetaan tinggi berdasarkan ukuran peta dan kaedah normalisasi. Persilangan data ditunjukkan dengan menggunakan kaedah analisis regresi dan hasil data yang diklasifikasikan semasa perintah kawalan pergerakan disahkan dalam bentuk peta pengorganisasian diri untuk mencapai objektif kajian. Dengan melakukan kaedah ini, kajian ini membuktikan bahawa hubungan antara permintaan tenaga terhadap hari darjah pendinginan, suhu purata, dan kes covid-19 sangat lemah dan pengesahan kaedah normalisasi *'logistic'* dan *'range'* menghasilkan hasil klasifikasi terbaik.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATION

C

Coefficient of determination
R²
.....
.....23
Cooling degree days
CDD 17
Corona Virus desease 2019
Covid-19..... 18

G

Grid system operator
GSO..... 26

K

Kementerian Kesihatan Malaysia
KKM 26
kiloWatt
kW 17

M

Matrix laboratory
MATLAB..... 23
MegaWatt
MW 15
Movement control order
MCO..... 25

S

Self-Organising Mapping
SOM 13

T

two-dimensional
2D..... 26

CHAPTER 1

INTRODUCTION

1.1 Background

It needs an effective method when dealing with large-scale data for a better future outcome and forecasting [1], mostly when data's growth rate was fast according to the recent trend for the past decades. In 2020, Tenaga Nasional Bhd. Malaysia is now developing a data analytics application for constructing a large solar (LLS) farm in Sepang to improve its operational efficiency [2]. This initiative is for the plan to reduce natural coal usage as a source of electricity generation. The energy demand in Malaysia increased in 2017 due to residential area development projects exceeding 64.4% [3]. This graph-based data classification will help the Malaysian government implement a roadmap to drive efficiency measures across all sectors to attain progressive energy savings.

Still, the challenges arise when the information has many factors to be considered [4], [5] to promote energy efficiency in households. However, to effectively gain insight into the future, the unsupervised method is required, such as data clustering. This study analysis is supported by using regression analysis and unsupervised method of self-organizing maps under normalization method. Data clustering provides patterns based on similarity for the unlabeled data called clusters [4], [6]. In 2013, T. Kohonen stated that the Self-organizing maps are widely recognized for clustering problems and data study in the industry [7].

In the power system application, the self-organizing maps are one of the best methods used in the classification process of fair and defective power distribution transformers [8], the research is using the numerical data for self-organizing maps to separate the defect transformer. This study will help understand the energy information pattern before and during the movement control order in Malaysia by using classification and clustering data analytics approach. The energy information-related data is obtained from different external sources and the Self-organizing maps is used because of its accuracy, producing excellent and well-ordered mapping quality.

1.2 Problem statements

- There are various data mining methods have improved on data analytics classification process. The classification method is needed, especially dealing with data from the past several years, which is time-consuming and complex to be performed by humans [4], [9].
- The classified data need to be analyzed to find the relevance impact on Energy Consumption in Malaysia with Cooling degree days, Daily Covid-19 Cases, and Average temperature in Malaysia.
- Based on the evidence from the references discovered, the self-organizing maps are applied in another area. Still, there is less research regarding data analytics related to energy consumption pattern [10].

1.3Objectives:

1) To investigate the data in finding significant connection between Energy consumption in Malaysia, Cooling degree days, Covid-19 Cases, and Average temperature in Malaysia using single Regression Analysis
2) To analyze data from different variable using multiple regression analysis.
3) To validate significant correlation using the best matching unit under the normalization method in the Self-Organising Mapping, SOM.

Table 1.1: Problem statement and objective mapping.

	Problem Statement 1	Problem Statement 2	Problem Statement 3
Objective 1	✓		
Objective 2		✓	
Objective 3		✓	✓

1.4 Scope

This study involves collecting the data from Grid System Operator to get energy consumption in Malaysia. The energy consumption will be analysed with three other variabels, which is Cooling degree days, Covid-19 Cases and Average temperature in Malaysia in order to observe the correlation and impact between them. The collected data will be processed while significant simulation on the classification and clustering algorithm would be applied simultaneously. Table 1.2 is the list of variables needed in this study.

Table 1.2: The data that will be collected.

Data	Unit
Power Demand	MW
Cooling Degree Days	Celcius
Covid-19 Case/Month	Person
Malaysia Average Temperature	Celcius

The required data for energy consumption is 2 years from year 2020 to 2021 in accordance with the before and during the Covid-19 pandemic making an appearence in Malaysia followed by 1 year of average temperature and cooling degree days. This study approach is using mathematical analysis and machine learning, which is Regression Analysis and clustering method. The clustering algorithm to be focused during the study completion is the self-organizing maps.

1.5 Flow of the Report

Chapter 1 and the rest will be devided into: 1.1 Background of the project, 1.2 Problem statement where the details of an issue to be addressed and a condition to be improved, 1.3 Objectives: The aim or target of the project,the scope of the study is defined in section 1.4 and 1.5 is indicaing the flow of the report.

The chapter 2 is started with 2.1 The introduction of the past research, 2.2 is where the information of the variables used in this study and divided into 4 subsection including, 2.2.1 explain about the energy demand in Peninsular Malaysia, 2.2.2 consisting of cooling degree days, 2.2.3 average temperature in Malaysia, 2.2.4 the final variable of covid-19 information. The chapter continues with section 2.3 where the energy-related custering algorithm method in previous study, which related to this study history.

The chapter 3 is mentioning on how the project is implemented, which started with 3.1 Methodology introduction, and section 3.2 explaining about data collection method, followed by mathematical approach 3.3 Regression analysis, the section 3.4 Constructing the self-organizing maps algorithm along with how the mapping analysis is done.

The chapter 4 is begin with 4.1 Introduction, and section 4.2 is discussion about the result in regression analysis, The energy demand MW is set as independant variable and the rest will be set as dependent variables, which constist of 4.2.1 Single Regression Analysis before the MCO, 4.2.2 Single Regression Analysis during the MCO, 4.3 is where the SOM results displayed including its subsection consisting of 4.3.1 Log Normalization method, 4.3.2 Logistic normalization method, 4.3.3 Range normalization method, 4.3.4 ‘Var’ normalization method, and lastly the chapter 4 is closed with 4.4 U-Matrix classification and 4.6 Sumary. Chapter 5 is divided into 2 section, which is 5.1 Conclusion of the study and 5.2 the future works of the study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter is going to approach past research that has been done by the researchers that are expert in electrical, data science, and machine computing. Therefore, the connected review of this field will focus on the research implemented in Malaysia and other countries regarding the similar method used in this study as well as other different approach.

In this chapter 2, section 2.2 contains energy information-related data that is consist of maximum demand, which represent the energy consumed measured in MW. Maximum demand increases every year along with the population as well as the sales of electricity in this country and it cost the energy provider planning the future installed capacity to avoid overload in some areas.

Different type of clustering method introduced in section 2.3. There are various type of way to conduct the data analysis, which are stated in the past research and comparing them for the best techniques and suitable method based on the research objectives. Different types of clustering techniques have been done to deal with data analytics from past papers and proved that every techniques have its own advantages and disadvantages affecting the result based on the researcher's requirements.

However, this study uses mathematical and neural network approach to analyze the data pattern, which is regression analysis in section 2.3.3 and SOM in 2.3.4. The energy information-related data in section 2.2 collected from diferent sources where it will undergo simulation on the classification and clustering algorithm to get the best mapping using the normalization method.

2.2 Information of Data Collection

2.2.1 Maximum Demand, MW

Maximum demand is the highest level of load in electrical order in a particular period of time, usually measured in kW. Figure 2.2 shows the maximum demand measured in MW since it involves large scale all over peninsular Malaysia covered by Tenaga Nasional Bhd with an average increment at a 1.8% rate every year [3]. It shows a good sign that Malaysia is growing since then, which energy is directly proportional to the in modern economic growth [11], positive economic growth will lead to an increase in energy demand.

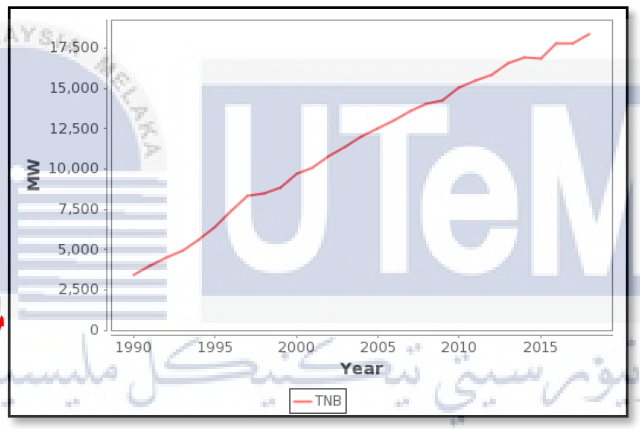


Figure 2.2: Maximum Demand *retrieved* from TNB

2.2.3 Cooling Degree Days

Electricity usage for residential costumers are highly depended on the outside temperature. Cooling requirement are highly dependent on the geographical region and needed when the average temperature is above 18ºC. In Malaysia, cooling degree days or CDD is the only one that can be counted because of its region is placed on the equator, which is 4.2105ºN, 101.9758ºE. It is defined that the CDD is the difference between the daily mean temperature and reference temperature [12].

In Equation (4), n is the number of days in a year, Ti is daily mean temperature for day i, Tb is the reference temperature for cooling. Cooling demand is influenced by factors such as air-conditioning purchasing capacity per capita and operating hours [13] and it means that the population can be a paramount factor.

2.2.4 Average Temperature

The climate in Malaysia is hot and humid. The Malaysian experience tropical weather year round because of its proximity to water, the climate is quiet humid. The average temperature in Malaysia is never too hot ranges between 29.79ºC to 33.50ºC [14]. The highest temperature ever recorded on 1988 was 40.1ºC in Perlis and another ocassion where the lowest temperature ever recorded was at Mount Kinabalu, Sabah at -2.4ºC on 2018 [15]. Figure 2.2 shows the average temperature in Malaysia taken from OGIMET and recorded that the average temperature in Malaysia is 25+ ºC to 26 ºC. The climate will affecting the energy consumption [16] and this study will find out the correlation between the temperature and energy consumption.

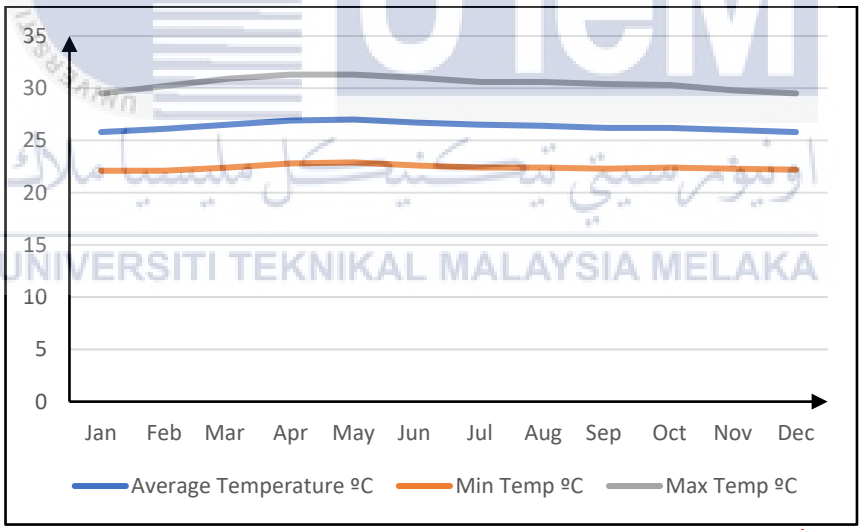


Figure 2.3: Malaysia Average Temperature

2.2.5 Covid-19 Cases

Coronavirus disease, also known as Covid-19, is an infectious disease that was discovered in December 2019 in Wuhan, Hubei [17]. The virus is spread by asymptomatic infected people and symptomatic people through oral fluid droplets, which are primarily airborne through coughing or sneezing [18]. In Malaysia, the first

cases was confirmed in Selangor where the person is recently returned from singapore [17]. Figure 2.3 show that monthly cases in Malaysia for 1 year period.

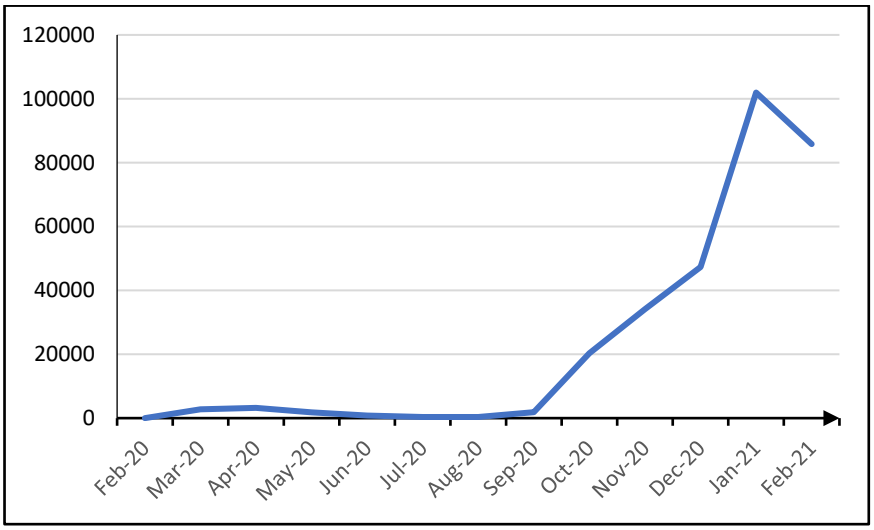


Figure 2.4: Monthly Cases

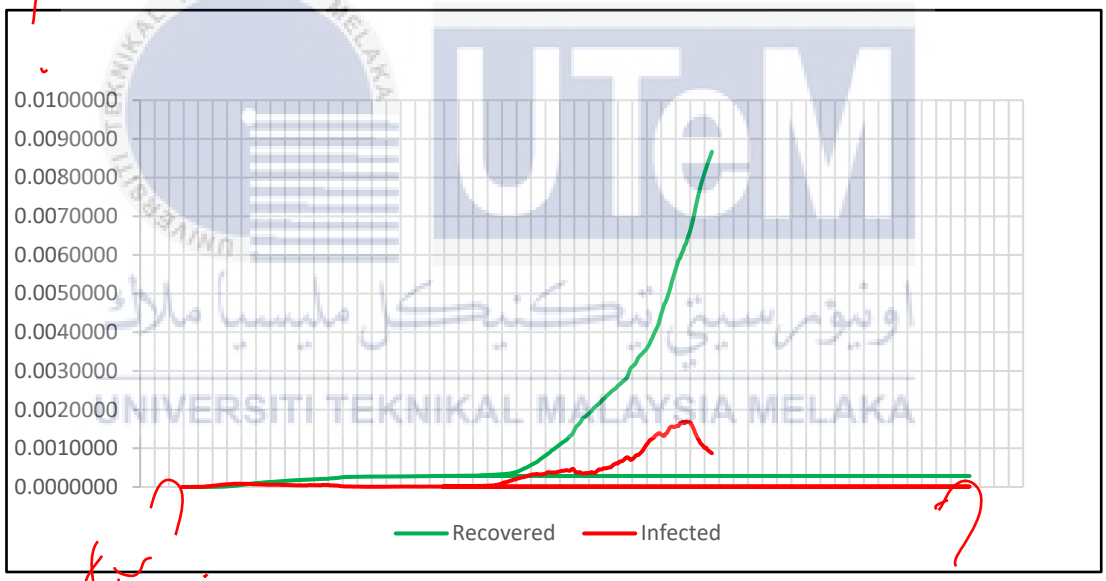


Figure 2.5: SEIR Model

Everyone over all ages is susceptible but the incubation period is depending on how strong the immunity of someone infected [18]. The data is taken from Kementerian Kesihatan Malaysia and the SEIR model stands for Susceptible, Exposed, Infectious, and recovered. In this model, the peak recovered in 0.0086 in 28 February 2021 and the infected peak was in 10 February 2021, which is 0.0016 over 31.53 million population in Malaysia. In this study, the analysis will be done to

investigate the connection between Covid-19 and Energy demand in Peninsular Malaysia using mathematical analysis and neural network classification.

2.3 Energy-Related Clustering Algorithm

Clustering methods are worn to identify groups having similarities in various data sets. Several clustering methods are used: K-means Algorithm, Hierarchical Algorithm, SOM algorithm.

2.3.1 K-means Algorithm

K-means is an algorithm clustering method which is commonly used by researcher in 1967 [19]. K-means algorithm is preferred when the size of the data sets is large. In 2018, China industry sectors using this method to assess their energy efficiency in every sector [20]; to define which company is having a problem in energy consumption and any environmental issues. The objects from different data sets are classified into one of the *k* groups before the centroid for each group is determined [21]. The data will be moved to the closest centroid to form a cluster and provide a visual map [22].

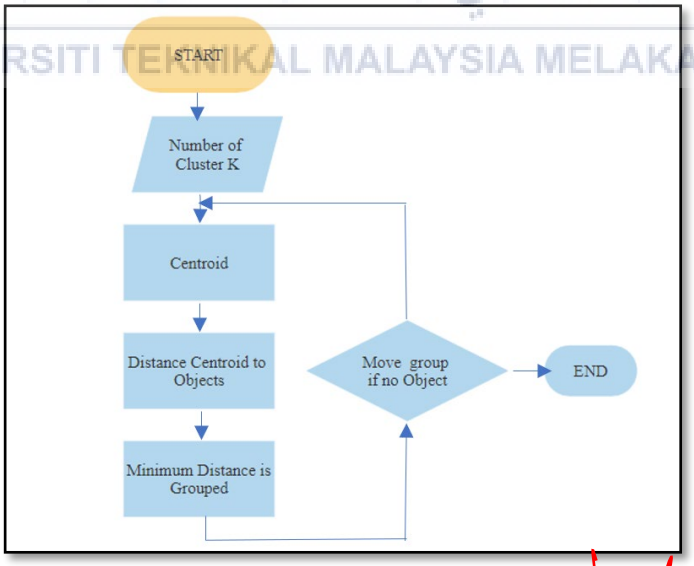


Figure 2.9: K-means flowchart

In this algorithm, data points are divided into k clusters according to their criterion for similitude calculation. *k*-means Algorithm has fast speed and a popular algorithm of clustering. The *k*-means algorithm is broadly used of some application such as vector quantization, cluster analysis, feature learning [23].

The algorithm is divided into two stages. The first step randomly selects *k* centres, where *k* is set in advance. The next stage is the object will be dragged to the nearest centroid after the Euclidean distance is calculated by using (1) and (2):

$$d(xi, yi) = \sum_{i=1}^{n} (xi - yi)^{1/2} \quad (1)$$

Where E is the summation of all error occur that is squared

$$E = \sum_{i=0}^{k} \sum_{x \in Ci} |x - xi|^2 \quad (2)$$

In shorts, *K*-means algorithm produces total of the squared interval from all samples in the cluster center minimum domain, which is dependent on the function of the criterion in the total of the squared error [24]. The iteration continues until the function is reduced, or data points are assigned stops interchange.

2.3.2 Hierarchical Algorithm

Hierarchical cluster analysis is an algorithm that assembles similar objects into groups called clusters, or it merely reapportions data groups to convergence [25]. Figure 2.10 shows the main result of this clustering method is a dendrogram, which shows the hierarchical relationship between clusters which reduces the routing complication [26].