EFFECTIVE VECTOR VARIANCE IN MODELING MALAYSIA HIGHWAY TRAFFIC NETWORK

NORHAIDAH MOHD ASRAH

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

AUGUST 2019

DEDICATION

This thesis is dedicated to my beloved family, thank you for the prayer, never ending support and encouragement for my success. To my friends who always be there when I need them.

ACKNOWLEDGEMENT

Alhamdulillah, praise to Allah for His blessings. With His consent, I would be able to complete my thesis. I would like to thank my supervisor, Prof Dr. Maman Abdurachman Djauhari on tutoring, guidance, encouragement, help, and patience during the preparation for my thesis. Thank you for your excellent concern and your valuable time spent for me during my research, to make sure that I am on the right track. I really appreciate your kindness not only to me but also to my family. He was my main supervisor from November 2011 until April 2014 when he was a contract Professor at Faculty of Science, Universiti Teknologi Malaysia (UTM). He was my co-supervisor when he joined Universitas Pasundan, Bandung, Indonesia from May 2014 to September 2014. Then, he joined the Institute for Mathematical Researcher (INSPEM) at Universiti Putra Malaysia (UPM) as Follow Researcher from September 2014 until 2015. Although he already left for Bandung, he still gives good commitment and continuous help to guide me to finish my thesis writing. I also want to thank AP Dr. Ismail Mohamad. He was my co-supervisor from May 2014 until the present. I really appreciate his opinions, suggestions, and help during my thesis writing. The comments, suggestions, ideas, and guides from both are really helping me in finishing my thesis. I also want to thank Universiti Tun Hussein Onn Malaysia (UTHM), UTM, UPM and Ministry of Higher Education for the opportunity to pursue my Ph.D., providing the facilities and the sponsorship. Lastly, thousands of thanks to Abah, Emak, Ahmad Wira (husband), Aamir Irfan (son), my family members and friends for their loves, encouragement and never-ending support during my study.

ABSTRACT

Highway traffic networks consist of in-coming vehicles to a toll plaza and out-coming vehicles from a toll plaza. The current practice to analyse a network is by using social network analysis (SNA) with the following three steps: (i) the network is considered as an undirected weighted complete graph (UWCG), (ii) the important information in the network is filtered using minimal spanning tree (MST), and (iii) the topological properties of each node is investigated using certain centrality measures. However, highway networks are complex and need a better method to analyse it. In this thesis, the Projek Lebuhraya Usahasama Berhad (PLUS) highway network is represented as a weighted directed network (WDN). The PLUS highway traffic network from 63 toll plazas in Peninsular Malaysia is studied to understand the in-coming and out-coming weights of traffic burden. Here, the weights of traffic burden refer to the in-coming and out-coming vehicles between two toll plazas. This study represents the complex PLUS highway traffic network as a non-symmetric matrix with positive finite element. The important information contain in the network are extracted using a unique and more robust method known as Forest. It is found that PLUS highway network produces only one MST in a network. This differs from other complex networks such as stock markets which typically produce more than one MST in a network. This study also investigates the dynamicity amongst PLUS highway toll plazas, which has not been discussed in other highway network studies. Using regression analysis on log return of traffic burden, it is found that Bukit Tambun Utara and Bukit Tambun Selatan are the two most dynamic toll plazas. In addition, the topological properties of network in this study use four types of centrality measures which are degree, betweenness, closeness and eigenvector. The performance of the toll plazas in the network can also be summarized based on an overall centrality measure using Principle Component Analysis. This approach is able to identify the most important toll plaza in the network. For instance, Sungei Besi toll plaza is found to be the most important toll plaza from the years 2009 until 2013 for in-coming traffic burden. However, this method could not identify the performance of the toll plazas based on the importance of their centrality measures. This is because it does not take into account, the multivariate dispersion of the centrality measures and thus cannot identify the most important centrality measure. An existing measure of multivariate dispersion is using effective variance (EV) based on the geometric mean of all eigenvalues. In this study, a new approach called effective vector variance (EVV) based on the arithmetic mean of all eigenvalues is used together with EV to identify the most important centrality measure. It is found that the most important centrality measure containing only one type of centrality measure is betweenness, the important centrality measure containing two types of centrality measures is betweenness and eigenvector, and the most important centrality measure containing three types of centrality measures is degree, betweenness and eigenvector. The results from this study can be used by the management of PLUS highway to improve its current system and operation.

ABSTRAK

Rangkaian trafik lebuhraya terdiri daripada kenderaan yang masuk ke plaza tol dan kenderaan yang keluar dari plaza tol. Kaedah semasa yang digunapakai untuk menganalisis rangkaian ialah dengan menggunakan analisis rangkaian sosial (SNA) dengan tiga langkah berikut: (i) rangkaian dianggap sebagai graf wajaran lengkap tak berarah (UWCG), (ii) maklumat penting dalam rangkaian ditapis dengan menggunakan pokok perentangan minimum (MST), dan (iii) sifat-sifat topologi setiap nod disiasat dengan menggunakan pengukuran pemusatan tertentu. Walau bagaimanapun, rangkaian lebuhraya adalah lebih rumit dan memerlukan kaedah yang lebih baik untuk menganalisisnya. Dalam tesis ini, rangkaian lebuhraya Projek Lebuhraya Usahasama Berhad (PLUS) telah diwakili sebagai rangkaian wajaran berarah (WDN). Rangkaian trafik lebuhraya PLUS daripada 63 buah plaza tol di Semenanjung Malaysia telah dikaji untuk memahami pemberat beban trafik yang masuk dan keluar. Di sini, pemberat kepada beban trafik merujuk kepada jumlah bilangan kenderaan yang masuk dan keluar di antara dua plaza tol. Kajian ini merujuk kepada rangkaian trafik PLUS yang kompleks sebagai matrik bukan simetrik dengan unsur terhingga yang positif. Maklumat penting yang terkandung dalam rangkaian ini diekstrak dengan menggunakan kaedah yang unik dan lebih teguh dikenali sebagai Forest. Didapati bahawa rangkaian lebuhraya PLUS menghasilkan hanya satu MST dalam rangkaiannya. Ini berbeza daripada rangkaian kompleks lain seperti pasaran saham yang biasanya menghasilkan lebih daripada satu MST dalam rangkaiannya. Kajian ini juga menyiasat dinamik di antara plaza-plaza tol di lebuhraya PLUS, yang belum pernah dibincangkan dalam kajian rangkaian lebuhraya yang lain sebelum ini. Dengan menggunakan analisis regresi pada log hasil pulangan bagi bebanan trafik, ia menunjukkan bahawa Bukit Tambun Utara dan Bukit Tambun Selatan adalah plaza tol yang paling dinamik. Tambahan lagi, sifat topologi dalam rangkaian kajian ini menggunakan empat jenis pengukuran pemusatan iaitu darjah, antara, kedekatan dan vektor eigen. Prestasi plaza tol dalam rangkaian juga boleh dirumuskan berdasarkan keseluruhan ukuran pemusatan menggunakan Analisis Komponen Utama. Pendekatan ini dapat mengenal pasti plaza tol yang paling penting dalam rangkaian. Sebagai contoh, plaza tol Sungei Besi adalah plaza tol yang paling penting daripada tahun 2009 hingga 2013 untuk beban lalu lintas yang masuk. Walau bagaimanapun, kaedah ini tidak dapat mengenal pasti prestasi plaza tol berdasarkan kepada kepentingan ukuran pemusatan. Ini kerana ia tidak mengambilkira penyebaran variasi berbilang bagi ukuran pemusatan dan seterusnya tidak dapat mengenal pasti ukuran pemusatan yang paling penting. Ukuran sedia ada bagi penyebaran variasi berbilang adalah dengan menggunakan varians yang berkesan (EV) berdasarkan pada purata geometrik untuk semua nilai eigen. Dalam kajian ini, satu pendekatan baru yang disebut sebagai varian vektor yang berkesan (EVV) yang berdasarkan pada nilai aritmetik untuk semua nilai eigen telah digunakan bersama dengan EV untuk mengenal pasti ukuran pemusatan yang paling penting. Kajian ini mendapati bahawa ukuran pemusatan yang paling penting yang mengandungi satu jenis ukuran pemusatan ialah antara, ukuran pemusatan yang paling penting yang mengandungi dua jenis ukuran pemusatan ialah antara dan vektor eigen, dan ukuran pemusatan yang paling penting yang mengandungi tiga jenis ukuran pemusatan ialah darjah, antara dan vektor eigen. Hasil daripada kajian ini boleh digunakan oleh pengurusan lebuhraya PLUS untuk memperbaiki sistem dan operasi mereka yang sedia ada.

TABLE OF CONTENTS

TITLE		PAGE
	DECLARATION	**ii**
	DEDICATION	**iii**
	ACKNOWLEDGEMENT	**iv**
	ABSTRACT	**v**
	ABSTRAK	**vi**
	TABLE OF CONTENTS	**vii**
	LIST OF TABLES	**xii**
	LIST OF FIGURES	**xv**
	LIST OF ABBREVIATIONS	**xix**
	LIST OF SYMBOLS	**xxii**
	LIST OF APPENDICES	**xxiii**
CHAPTER 1	**INTRODUCTION**	**1**
1.1	Introduction	1
1.2	Background of the Problem	1
1.3	Problem Statement	4
1.4	Research Objectives	5
1.5	The Scope of the Study	6
1.6	Research Data	7
1.7	The Significance of the Study	7
1.8	Outline of the Thesis	8
1.9	Conclusion	9
CHAPTER 2	**LITERATURE REVIEW**	**11**
2.1	Introduction	11
2.2	Minimal Spanning Tree (MST)	11
2.3	Kruskal’s Algorithm and Prim’s	13
2.4	*Forest* of All MSTs	16

2.5	Social Network Analysis		17
2.6	Centrality Measures		19
2.7	Multivariate Dispersion Measure		23
2.8	Conclusion		25
CHAPTER 3	**RESEARCH METHODOLOGY**		**27**
3.1	Introduction		27
3.2	Data Collection		27
3.3	Representation of PLUS Highway Traffic Network		29
	3.3.1	Graph representation	29
	3.3.2	Adjacency matrix representation	29
3.4	The Most Dynamic Toll Plazas		30
3.5	Information Filtering Based on MST		32
3.6	*Forest* Based Filtering		33
3.7	Filtered Network Analysis		39
	3.7.1	Degree centrality measure	39
	3.7.2	Betweenness centrality measure	39
	3.7.3	Closeness centrality measure	40
	3.7.4	Eigenvector centrality measure	40
	3.7.5	Overall centrality measure	41
3.8	The Most Important Centrality Measure		42
	3.8.1	Effective variance (EV)	43
	3.8.2	Effective vector variance (EVV)	46
3.9	The Research Framework		49
3.10	Conclusion		51
CHAPTER 4	**RESULTS ON IN-COMING TRAFFIC BURDEN**		**53**
4.1	Introduction		53
4.2	PLUS Highway Network as Complex Network		54
4.3	The Dynamicity of Toll Plazas		56
4.4	Network Filtering		65
	4.4.1	In-coming traffic burden filtering network for 2009	66

4.4.2 In-coming traffic burden filtering network for 2010 69
4.4.3 In-coming traffic burden filtering network for 2011 71
4.4.4 In-coming traffic burden filtering network for 2012 73
4.4.5 In-coming traffic burden filtering network for 2013 74
4.5 Evidence from Centrality Measures 76
4.5.1 Degree centrality measure 76
4.5.1.1 In-coming traffic burden degree centrality for 2009 until 2013 77
4.5.2 Betweenness centrality measure 80
4.5.2.1 In-coming traffic burden betweenness centrality for 2009 until 2013 80
4.5.3 Closeness centrality measure 83
4.5.3.1 In-coming traffic burden closeness centrality for 2009 until 2013 83
4.5.4 Eigenvector centrality measure 86
4.5.4.1 In-coming traffic burden eigenvector centrality for 2009 until 2013 86
4.5.5 Overall Centrality Measure 88
4.5.5.1 In-coming traffic burden overall centrality for 2009 until 2013 89
4.6 The Most Important Centrality Measure 91
4.6.1 Yearly covariance matrix 91
4.6.2 Yearly generalized variance (GV) and vector variance (VV) for in-coming traffic burden 93
4.6.3 The most important measure of effective variance (EV) viewpoint 94
4.6.4 The most important measure of effective vector variance (EVV) viewpoint 95
4.7 The Most Important Two Centrality Measures 96
4.7.1 The most important two centrality measures relative to EV 96

4.7.2 The most important two centrality measures relative to EVV 97

4.8 The Most Important Three Centrality Measures 101

4.8.1 The most important three centrality measures relative to EV 101

4.8.2 The most important three centrality measures relative to EVV 102

4.9 Conclusion 106

CHAPTER 5 RESULTS ON OUT-COMING TRAFFIC BURDEN 109

5.1 Introduction 109

5.2 Out-Coming Traffic Burden Network 109

5.3 Network Filtering 118

5.3.1 Out-coming traffic burden filtering network for 2009 119

5.3.2 Out-coming traffic burden filtering network for 2010 121

5.3.3 Out-coming traffic burden filtering network for 2011 124

5.3.4 Out-coming traffic burden filtering network for 2012 126

5.3.5 Out-coming traffic burden filtering network for 2013 127

5.4 Evidence from Centrality Measures 129

5.4.1 Degree centrality measure 129

5.4.1.1 Out-coming traffic burden degree centrality for 2009 until 2013 130

5.4.2 Betweenness centrality measure 132

5.4.2.1 Out-coming traffic burden betweenness centrality for 2009 until 2013 132

5.4.3 Closeness centrality measure 135

5.4.3.1 Out-coming traffic burden closeness centrality for 2009 until 2013 135

5.4.4 Eigenvector centrality measure 137

5.4.4.1 Out-coming traffic burden eigenvector centrality for 2009 until 2013 137

5.4.5 Overall centrality measure 140

5.4.5.1 Out-coming traffic burden overall centrality for 2009 until 2013 140

5.5 The Most Important Centrality Measure 142

5.5.1 Yearly covariance matrix 143

5.5.2 Yearly generalized variance and vector variance 144

5.5.3 The most important measure from effective variance viewpoint 145

5.5.4 The most important measure from effective vector variance viewpoint 145

5.6 The Most Important Two Centrality Measures 147

5.6.1 The most important two centrality measures relative to EV 147

5.6.2 The most important two centrality measures relative to EVV 148

5.7 The Most Important Three Centrality Measures 152

5.7.1 The most important three centrality measures relative to EV 152

5.7.2 The most important three centrality measures relative to EVV 152

5.8 Conclusion 157

CHAPTER 6 CONCLUSIONS 159

6.1 Introduction 159

6.2 Mathematical Areas 159

6.3 Physical Research Areas 162

6.4 Conclusions 165

6.5 Future Research 166

REFERENCES 167

LIST OF PUBLICATIONS 180

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Ratio of computation speed of CD to VV	25
Table 4.1	Colour of the cell in the dot plot matrix	57
Table 4.2	The monthly average of in-coming traffic burden at each toll plaza	61
Table 4.3	Parameter estimates and autocorrelation of in-coming traffic burden log return (Ordered by autocorrelation).	64
Table 4.4	Colour of the cell in dot plot matrix representing *Forest*	66
Table 4.5	Colour of the nodes for *Forest*	66
Table 4.6	Degree centrality of in-coming traffic burden for 2009 until 2013	78
Table 4.7	Betweenness centrality of in-coming traffic burden for 2009 until 2013	81
Table 4.8	Closeness centrality of in-coming traffic burden for 2009 until 2013	84
Table 4.9	Eigenvector centrality of in-coming traffic burden for 2009 until 2013	87
Table 4.10	Overall centrality of in-coming traffic burden for 2009	89
Table 4.11	Covariance matrix of the four centrality measures for in-coming traffic burden for 2009.	92
Table 4.12	Covariance matrix of the four centrality measures for in-coming traffic burden for 2010.	92
Table 4.13	Covariance matrix of the four centrality measures for in-coming traffic burden for 2011.	92
Table 4.14	Covariance matrix of the four centrality measures for in-coming traffic burden for 2012.	92
Table 4.15	Covariance matrix of the four centrality measures for in-coming traffic burden for 2013.	93
Table 4.16	Covariance matrix of the four centrality measures for in-coming traffic burden for five years (2009 to 2013).	93
Table 4.17	Yearly GV and VV for in-coming traffic burden	94
Table 4.18	The yearly ratio of each centrality measures relative to EV for in-coming traffic burden	94

Table 4.19 The yearly ratio of each measure relative to EVV for in-coming traffic burden 95

Table 4.20 Yearly most important measures according to EV and EVV for in-coming traffic burden 96

Table 4.21 The most important two centrality measures relative to EV for in-coming traffic burden 97

Table 4.22 The most important pair relative to EVV for in-coming traffic burden 97

Table 4.23 The most important three centrality measures relative to EV for in-coming traffic burden 102

Table 4.24 The most important three centrality measures relative to EVV for in-coming traffic burden 102

Table 4.25 Summary of the most important centrality measures for in-coming traffic burden 106

Table 5.1 Color of cell in dot plot matrix 110

Table 5.2 The monthly average of out-coming traffic burdens at each toll plaza 113

Table 5.3 Parameter estimates and autocorrelation of out-coming traffic burden log return (Ordered by autocorrelation) 116

Table 5.4 Colour of the cell in dot plot matrix representing *Forest*. 118

Table 5.5 Colour of nodes for *Forest.* 119

Table 5.6 Degree centrality of out-coming traffic burden for 2009 until 2013 130

Table 5.7 Betweenness centrality of out-coming traffic burden for 2009 until 2013 133

Table 5.8 Closeness centrality of out-coming traffic burden for 2009 until 2013 135

Table 5.9 Eigenvector centrality of out-coming traffic burden for 2009 until 2013 138

Table 5.10 Overall centrality of out-coming traffic burden for 2009 until 2013 140

Table 5.11 Covariance matrix of the four centrality measures for out-coming traffic burden for 2009. 143

Table 5.12 Covariance matrix of the four centrality measures for out-coming traffic burden for 2010. 143

Table 5.13 Covariance matrix of the four centrality measures for out-coming traffic burden for 2011. 143

Table 5.14 Covariance matrix of the four centrality measures for out-coming traffic burden for 2012. 143

Table 5.15 Covariance matrix of the four centrality measures for out-coming traffic burden for 2013. 144

Table 5.16 Covariance matrix of the four centrality measures for out-coming traffic burden for five years (2009 to 2013). 144

Table 5.17 Yearly GV and VV for out-coming traffic burden. 144

Table 5.18 Yearly EV for out-coming traffic burden. 145

Table 5.19 Yearly EVV for out-coming traffic burden. 146

Table 5.20 Yearly most important measures according to EV and EVV for out-coming traffic burden. 146

Table 5.21 The most important pair relative to EV for out-coming traffic burden. 147

Table 5.22 The most important two centrality measures relative to EVV for out-coming traffic burden 148

Table 5.23 The most important three centrality measures relative to EV for out-coming traffic burden. 152

Table 5.24 The most important three centrality measures relative to EVV for out-coming traffic burden. 153

Table 5.25 Summary of the most important centrality measures for out-coming traffic burden 157

Table 6.1 The summary for mathematical areas scope 162

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	Complex network with three nodes A, B, and C	15
Figure 2.2	The two possible MSTs	15
Figure 2.3	Example of a directed graph (a) and undirected graph (b)	18
Figure 2.4	Example of degree centrality measure	20
Figure 2.5	Example of betweenness centrality measure	20
Figure 2.6	Example of closeness centrality measure	21
Figure 2.7	Example of eigenvector centrality measure	21
Figure 3.1	The example of data provided by PLUS highway management	28
Figure 3.2	The flow chart to construct SDU and *Forest*	37
Figure 3.3	The information extraction based on MST and *Forest*	38
Figure 3.4	The research framework	50
Figure 4.1	PLUS highway in Peninsular Malaysia	55
Figure 4.2	Graph representation of traffic network on PLUS highway in 2013	56
Figure 4.3	Dot plot matrix of in-coming traffic burden for 2009	58
Figure 4.4	Dot plot matrix of in-coming traffic burden for 2010	58
Figure 4.5	Dot plot matrix of in-coming traffic burden for 2011	59
Figure 4.6	Dot plot matrix of in-coming traffic burden for 2012	60
Figure 4.7	Dot plot matrix of in-coming traffic burden for 2013	60
Figure 4.8	The total monthly average of in-coming traffic burden.	63
Figure 4.9	Dot plot matrix representing the *Forest* for 2009 in-coming traffic burden	67
Figure 4.10	In-coming traffic burden *Forest* for 2009	68
Figure 4.11	Dot plot matrix representing the *Forest* for 2010 in-coming traffic burden	70
Figure 4.12	In-coming traffic burden *Forest* for 2010	71
Figure 4.13	Dot plot matrix representing the *Forest* for 2011 in-coming traffic burden	72

Figure 4.14 In-coming traffic burden *Forest* for 2011 72

Figure 4.15 Dot plot matrix representing the *Forest* for 2012 in-coming traffic burden 73

Figure 4.16 In-coming traffic burden *Forest* for 2012 74

Figure 4.17 Dot plot matrix representing the *Forest* for 2013 in-coming traffic burden 75

Figure 4.18 In-coming traffic burden *Forest* for 2013 75

Figure 4.19 Graphical representation of the importance level between the two centrality measures for EV and EVV for 2009 in-coming traffic burden 98

Figure 4.20 Graphical representation of the importance level between the two centrality measures for EV and EVV for 2010 in-coming traffic burden 98

Figure 4.21 Graphical representation of the importance level between the two centrality measures for EV and EVV for 2011in-coming traffic burden 99

Figure 4.22 Graphical representation of the importance level between the two centrality measures for EV and EVV for 2012 in-coming traffic burden 99

Figure 4.23 Graphical representation of the importance level between the two centrality measures for EV and EVV for 2013 in-coming traffic burden 100

Figure 4.24 Graphical representation of the importance level between the two centrality measures for EV and EVV for five years in-coming traffic burden 100

Figure 4.25 Graphical representation of the importance level between three centrality measures of EV and EVV for 2009 in-coming traffic burden 103

Figure 4.26 Graphical representation of the importance level between three centrality measures of EV and EVV for 2010 in-coming traffic burden 103

Figure 4.27 Graphical representation of the importance level between three centrality measures of EV and EVV for 2011 in-coming traffic burden 104

Figure 4.28 Graphical representation of the importance level between three centrality measures of EV and EVV for 2012 in-coming traffic burden 104

Figure 4.29 Graphical representation of the importance level between three centrality measures of EV and EVV for 2013 in-coming traffic burden 105

Figure 4.30 Graphical representation of the importance level between three centrality measures of EV and EVV for 5 years in-coming traffic burden 105

Figure 5.1 Dot plot matrix of out-coming traffic burden network for 2009. 110

Figure 5.2 Dot plot matrix of out-coming traffic burden network for 2010. 111

Figure 5.3 Dot plot matrix of out-coming traffic burden network for 2011. 112

Figure 5.4 Dot plot matrix of out-coming traffic burden network for 2012. 112

Figure 5.5 Dot plot matrix of out-coming traffic burden network for 2013. 113

Figure 5.6 The total monthly average of out-coming traffic burden. 116

Figure 5.7 Dot plot matrix representing the *Forest* for 2009 out-coming traffic burden. 120

Figure 5.8 Out-coming traffic burden *Forest* for 2009 120

Figure 5.9 Dot plot matrix representing the *Forest* for 2010 out-coming traffic burden. 123

Figure 5.10 Out-coming traffic burden *Forest* for 2010 123

Figure 5.11 Dot plot matrix representing the Forest for 2011 out-coming traffic burden. 125

Figure 5.12 Out-coming traffic burden *Forest* for 2011 125

Figure 5.13 Dot plot matrix representing the *Forest* for 2012 out-coming traffic burden. 126

Figure 5.14 Out-coming traffic burden *Forest* for 2012 127

Figure 5.15 Dot plot matrix representing the *Forest* for 2013 out-coming traffic burden 128

Figure 5.16 Out-coming traffic burden *Forest* for 2013 128

Figure 5.17 Graphical representation of the importance level between two centrality measures for EV and EVV for 2009 out-coming traffic burden. 149

Figure 5.18 Graphical representation of the importance level between two centrality measures for EV and EVV for 2010 out-coming traffic burden. 149

Figure 5.19 Graphical representation of the importance level between two centrality measures for EV and EVV for 2011 out-coming traffic burden. 150

Figure 5.20 Graphical representation of the importance level between two centrality measures for EV and EVV for 2012 out-coming traffic burden. 150

Figure 5.21 Graphical representation of the importance level between two centrality measures for EV and EVV for 2013 out-coming traffic burden. 151

Figure 5.22 Graphical representation of the importance level between two centrality measures for EV and EVV for five years out-coming traffic burden. 151

Figure 5.23 Graphical representation of the importance level between three centrality measures for EV and EVV for 2009 out-coming traffic burden. 154

Figure 5.24 Graphical representation of the importance level between three centrality measures for EV and EVV for 2010 out-coming traffic burden. 154

Figure 5.25 Graphical representation of the importance level between three centrality measures for EV and EVV for 2011 out-coming traffic burden. 155

Figure 5.26 Graphical representation of the importance level between three centrality measures for EV and EVV for 2012 out-coming traffic burden. 155

Figure 5.27 Graphical representation of the importance level between three centrality measures for EV and EVV for 2013 out-coming traffic burden. 156

Figure 5.28 Graphical representation of the importance level between three centrality measures for EV and EVV for five years out-coming traffic burden. 156

LIST OF ABBREVIATIONS

EV	-	Effective Variance
EVV	-	Effective Vector Variance
GV	-	Generalized Variance
MST	-	Minimum Spanning Tree
PLUS	-	Projek Lebuhraya Usahasama Berhad
SDU	-	Subdominant Ultrametric
SNA	-	Social Network Analysis
VV	-	Vector Variance
WDG	-	Weighted Directed Graph
JRU	-	Juru
BTU	-	Bukit Tambun (U)
BTS	-	Bukit Tambun (S)
JWI	-	Jawi
BBR	-	Bandar Baru
BKM	-	Bukit Merah
TPU	-	Taiping (U)
CKJ	-	Changkat Jering
KKS	-	Kuala Kangsar
IPU	-	Ipoh (U)
IPS	-	Ipoh (S)
SPP	-	Simpang Pulai
GPG	-	Gopeng
TPH	-	Tapah
BDR	-	Bidor
SKI	-	Sungkai
PSR	-	Persimpangan Slim River
BRG	-	Behrang
TGM	-	Tanjung Malim
LBB	-	Lembah Beringin
BTR	-	Bukit Tagar

BKB	-	Bukit Beruntung
RAW	-	Rawang
RWS	-	Rawang Selatan
HSB	-	Sungai Buloh (U)
SGB	-	Sungai Buloh
JLD	-	Jalan Duta
KDR	-	Kota Damansara
DMR	-	Damansara
SBG	-	Subang
STA	-	Setia Alam
BKR	-	Bukit Raja
SBI	-	Sungei Besi
UPM	-	UPM
KJG	-	Kajang
BGS	-	Bangi
PPM	-	Persimpangan Putra Mahkota
NLI	-	Nilai
SHA	-	Shah Alam
EBN	-	Ebor (N)
EBS	-	Ebor (S)
SEA	-	Seafield
USJ	-	USJ
PHT	-	Putra Height
KLA	-	KLIA
SBN	-	Seremban
PDU	-	Port Dickson (U)
PDS	-	Port Dickson (S)
SWG	-	Senawang
PLI	-	Pedas Linggi
SAT	-	Simpang Ampat
AKH	-	Ayer Keroh
JSN	-	Jasin
TGK	-	Tangkak

PGH	-	Pagoh
YPU	-	Yong Peng (U)
YPS	-	Yong Peng (S)
AHT	-	Ayer Hitam
MAC	-	Machap
SPR	-	Simpang Renggam
SDK	-	Sedenak
KLI	-	Kulai
SNU	-	Senai (U)
SKD	-	Skudai

LIST OF SYMBOLS

λ	-	Eigenvalues
D, d	-	Distance matrix, the elements of D
Tr	-	Trace
R^2	-	Correlation
$\mathbf{A}$	-	Adjacency matrix
Σ	-	Covariance matrix

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	The Dynamicity of a Toll Plaza Computation Using Microsoft Excel	174
Appendix B	EV and EVV Computation Using Microsoft Excel	175

CHAPTER 1

INTRODUCTION

1.1 Introduction

Network analysis can be found in many fields such as medical, organizations, economics, ecologies, technologies, transportations and many more. The relationships in the network explain the systems which contain each element that is connected among them. These connections are the topics of interest and the network topology is investigated. The information contains in the network can be filtered and interpret with the application of social network analysis (SNA). The visualization of each network can be presented by using the theoretical graph approach while the important information in the network can be extracted by using MST and *Forest*. The behavior of each centrality measures is one of the topic interests. The most important centrality measures can be defined by looking at each dispersion measure.

1.2 Background of the Problem

The highway is one of the most important facilities to transfer passengers as well as consumer goods. Therefore, the highway has the potential to support economic growth. It contributes directly to socio-economic development and thus improves the quality of life. Industrialization and economic growth have been rapidly increasing once Projek Lebuhraya Usahasama Berhad (PLUS) highway network in Malaysia was built. Hence, the patterns of transportation also changed following the economic development. The growth of the Malaysian highway network is the result and the cause of this positive progress of industrialization. The demand for traveling increases as the need for people mobility increases. It is also related and associated with the demands and supplies in manufacturing sectors, rapid urban development, and population growth.

Malaysia is highly committed to building the development of road network. According to the Department of Statistics, Malaysia (DOSM, 2016), the road mileage in Malaysia increased from 214,813 km (2015) to 236,802 km (2016). The increase in road mileage is 9.29% in a year. Meanwhile, the number of registered motor vehicles also increases from 26,301,952 (2015) to 27,613,259 (2016) marking an increase of about 4.75% in a year. These data show that Malaysia needs good road networks. The registered motor vehicles here refer to motorcycle, car, taxi and hired car, bus, consumer goods vehicle and truck. At the same time, Johor Bahru – Kulai station (JR 204) is the highest traffic composition with 110,498 vehicles in 2016 (MOT, 2016). These vehicles include cars, taxis, lorries, buses, and motorcycles.

To the knowledge of the author, there is a limited number of published works on highway network analysis. The published works include Messmer et al. (1998), Gabor and Csabai (2002), Sohn (2006), Villas et al. (2009), Bounova (2009), Park and Yilmaz (2010), and Haibin (2016). These works use the theoretical graph approach. Messmer et al. (1998) developed the Variable Message Sign (VMS); an information and guidance system in the interurban Scottish highway network. In 2002, Gabor and Csabai (2002) studied the comparison of the characteristic between highway network and computer network traffic in a simple topology. Boccaletti et al. (2006) defined it as networks in which a real number is connected to every link. Sohn (2006) studied the accessibility of the highway network in Maryland, USA, under flood. Villas et al. (2009) construct a model for different worldwide highway networks by using Geographic Path Network (GPN).

Bounova (2009) studied the topological network evolution for US Airline and the growth of language Wikipedia network. Park and Yilmaz (2010) studied road network analysis in Columbus Ohio and Washington D.C by using centrality measure. Meanwhile, Haibin (2016) applied the complex network theory to analyze the statistical features of the Guangdong Province Highway Network (GDHN) in China.

Theoretical graph approach is selected because it can visualize the network easily and able to analyze the highway network more comfortably. In graph theory, a graph represents a set consisting of objects (also called nodes) and link (or, edge) connecting two nodes. The edge may be directed or undirected. When the edges have no orientation, the graph is called 'undirected graph' while the graph with direction is called 'directed graph' (Bondy and Murty, 1976).

In this thesis, the highway network is considered as a weighted directed graph (WDG). It is a graph where each link has its own weight. Boccaletti et al. (2006) defined it as networks in which a real number is connected to every link. Here, a node represents a toll plaza and the link is the highway connecting them. Traditionally, in terms of geographical location, the weight of the link between two toll plazas is the length of highway (in mile or kilometer) between them. However, in economic terms, the weight of link could be the length of time from one toll plaza to the other, traffic burden, cost and weight of transported goods.

The weight of the link in a network can be described as a network with the connections between nodes that have weights or burdens allocated to them. Since the above-mentioned published works do not use traffic burden to define the weight of the link, in this thesis it is used as 'the number of vehicles in-coming and out-coming through toll plaza'.

Once the highway network is represented by a weighted directed graph (WDG), it becomes a complex network. The problem is on how to filter the information therein, analyze the filtered network, identify the toll plazas that must be paid more attention to and determine the most important centrality measure that can be used by PLUS highway management to develop highway network.

The important information contains in the network can be filtered by applying the minimal spanning tree (MST) and *Forest*. MST is used in many fields for filtering or extracting the important information in a complex network (Mantegna (1999), Eom et al. (2009) and Tabak et al. (2010)). Although MST can extract the important information from the complex network, it is not unique and robust. To handle these weaknesses, Djauhari (2012) introduced *Forest* to avoid the producing of misleading information from MST.

Centrality measures are performed to learn the topological characteristic of the network. The four types of centrality measures that are usually applied by the researchers are degree, betweenness, closeness and eigenvector centrality measures. The research on degree centrality measure can be found in Freeman (1978), Latora and Marchiori (2007) and Chen et al. (2012). Meanwhile, the betweenness centrality measure can be found in the research by Borgatti (1995 and 2005) and Opsahl et al. (2010). Furthermore, the research on closeness centrality measure can be found in Newman (2008), Crucitti et al. (2006) and Brandess and Fleischer (2005). Other than that, research by Borgatti (2005) and Bonacich (1972) are examples of research done on eigenvector centrality measure. Based on these centrality measures, it is crucial to determine which centrality measure is the most important centrality measure.

1.3 Problem Statement

The complex network that represents the PLUS highway traffic burden consists of important information of the in-coming and out-coming traffic burden. The important information covers the whole toll plazas along the PLUS highway. According to the graph theory, the current practice of complex network analysis is usually conducted by considering the network as a directed or an undirected weighted complete graph. In this study, the network is considered a weighted directed graph (WDG). The edge weight is known as traffic burden which is measured by the number of vehicles coming-in to a toll plaza and coming-out from a toll plaza. Thus, the main problems encountered in this thesis consist of how to analyze highway network from a complex network analysis point of view.

The social network analysis is one of the methods to understand the information contained in the highway network. By using graph theory, it visualized the network in WDG. Meanwhile, in algorithm theory, the network is represented as a non-symmetric matrix. Based on the matrix, the next problem encountered in this thesis is how to analyze the matrix so that the most important information can be extracted?

Based on the previous research, MST is applied to analyze the non-symmetric matrix. The weakness of using MST is that it is not unique and not robust. It is not unique because it has more than one MST in a network. The result can be confusing since it cannot give the most important answer. While in terms of interpretation, MST is not robust since the interpretation depends on many MST in the network. The weakness of MST could lead to misleading information. To overcome the weakness of MST, a method that can produce a unique and robust result is needed.

The next step in analyzing the network is to analyze the topological properties of the network. Based on the four common centrality measures, the most important centrality measure is proposed. Among the centrality measures, there is one measure that qualities as the most important centrality measure. The effective variance (EV) is one of the existing methods in multivariate dispersion that can be used to find the most important centrality measure. However, EV has its own limitations. A new alternative or method should be proposed to handle the limitations of EV. Thus, the new method is proposed in this thesis.

In addition, the dynamicity of the toll plaza is also drawn attention in the thesis since the dynamicity of a node in the network has never been discussed before in previous studies. The main interest here is to know whether the most important toll plaza is also the dynamic toll plaza.

1.4 Research Objectives

To overcome the problems mentioned above, the main objectives of this study are:

(i) to model the PLUS highway network as a non-symmetric matrix with the positive finite element and/or infinite element.

(ii) to measure the dynamicity of a toll plaza by using autocorrelation, in terms of traffic burden log return (the log return of traffic burden data).

(iii) to analyze the *Forest* by using the traditional four centrality measures namely degree, betweenness, closeness, and eigenvector. Then, the overall centrality measure is defined by using the Principal Component Analysis (PCA).

(iv) to determine the most important centrality measure in term of the maximum variance.

1.5 The Scope of the Study

The scope of this study can be divided into three aspects. The three aspects are:

(i) Data

The data used in this study are limited to the data prepared by Toll Department of PLUS Malaysia Sdn. Bhd. The number of in-coming and out-coming traffic burden or the number of vehicles is collected every month from July 2009 until December 2013. These traffic burdens are recorded from 65 toll plazas along PLUS highway, starting from Juru until Skudai. All data are fully used for the analysis. The data is free from missing values, latest (at that time) and relevant for this study.

(ii) Methodology

The methodologies used in this study are referred to the scope of the mathematical areas. The first stage is to visualize the highway network. This stage needs the use of graph theory to visualize the network based on the WDG. Then, the extraction of important information from the network can be determined using MST and *Forest*. The dynamicity of the toll plaza in the network can be recognize using the regression on the algorithmic returns. Furthermore, the network topology of the highway network can be interpreted by applying the four most centrality measures.

Lastly, the most important centrality measure can be found by using EV and EVV. These methods are based on the multivariate dispersion measure.

The effective variance is defined based on generalized variance (GV) while EVV is defined based on vector variance (VV). EV is the geometric mean of all eigenvalues while EVV is the arithmetic mean of all squared eigenvalues. Since EV have the limitations in measuring the dispersion, EVV is introduced to manage and handle the limitations of EV.

(iii) Application

The application of the methodologies mentioned in (ii) covers the scope for physical research areas. The physical research area is the PLUS highway traffic network which consists of 65 toll plazas spread out in eight states (Penang, Perak, Kedah, Selangor, Wilayah Persekutuan, Melaka, Negeri Sembilan, and Johor).

1.6 Research Data

The data involved in this study is about the number of vehicles (traffic burden) in-coming to a toll plaza and out-coming from a toll plaza. The data collected from 63 toll plazas along PLUS highway. The toll plazas are from Penang (3), Kedah (1), Perak (14), Selangor (25), Wilayah Persekutuan (3), Negeri Sembilan (5), Melaka (3) and Johor (11).

1.7 The Significance of the Study

The significance of the study can be divided into two different categories. The first category will refer to the development of mathematical areas. Other than MST and *Forest* for extracting the important information in the complex network, the dynamicity of the toll plaza is a highlight as one of the contributions in this study. However, the foremost contribution of this study is EVV. EVV is introduced to identify the most important centrality measures in terms of multivariate dispersion.

The second category refers to the management of physical research areas that is PLUS highway traffic network. It is hoped that PLUS highway management can get many advantages and benefits from this thesis. Starting from the visualization of the highway traffic network from 2009 until 2013, it can help PLUS highway management to visualize the connection of toll plaza with other toll plazas. Then, the information filtering based on *Forest* and centrality measures, it will help them to determine which toll plazas are important with the function given by each centrality measures. With the help of EV and EVV, PLUS highway management can determine which toll plaza with the most important centrality measure, the most important two centrality measures, and the most important three centrality measures.

1.8 Outline of the Thesis

This thesis is organized into six chapters. Chapter 1 is the Introduction. This chapter will explain the background of the problem, problem statement, research objectives, scopes of the study, chapter outline, the significance of the study and conclusion.

In Chapter 2, discussion on the previous research about the MST, *Forest* of all MSTs, SNA analysis, in brief, centrality measures, multivariate dispersion measures, and conclusions are presented and described in detail. Based on these discussions, the proposed methods can be developed by fulfilling the gap from the previous research.

Chapter 3 in this thesis explain about the methodologies of this research. It started with an explanation about data collection. Then, the representation of the PLUS highway traffic network is discussed. It related to the graph and the adjacency matrix representation. It continues with the discussion on the most dynamic toll plazas. After that, the information filtering based on MST and *Forest* is discussed. Next, the filtered network analysis based on the centrality measures namely degree, betweenness, closeness, eigenvector, and overall centrality measures are presented according to their methods. It continues with the most important centrality measure by using EV and EVV.

In Chapter 4 and 5, the results on in-coming traffic burden and out-coming traffic burden are discussed in detail. It includes the dynamicity of toll plazas, the network filtering, the evidence from centrality measures based on the degree, betweenness, closeness, eigenvector, and overall centrality measures. Lastly, the most important centrality measures are discussed according to EV and EVV.

In the last Chapter 6, the conclusion of the research scopes is discussed. It divided into two topics, the mathematical areas, and the physical research areas. The recommendations for PLUS highway management are also highlighted in this chapter.

1.9 Conclusion

This chapter explained the background and ideas of this study. The background of the study explained how the study is developed from the beginning. Then, the problem statement defines the lack and gap of the study. It then continues with the research objectives to solve the problem arise before. The study is restricted to the scope given that covers the mathematical areas and physical research area. To gain more information about the previous studies based on the case studies and methodologies, it discussed comprehensively in Chapter 2.

REFERENCES

Alexander, K. (1995) ‘Percolation and minimal spanning forests in infinite graphs’, *The Annals of Probability*, 23(1), 87-104.

Alt, F.B. and Smith, N.D. (1988) *Multivariate process control*, in Krishnaiah, P.R., Rao, C. R. (eds.) *Handbook of Statistics*. Elsevier Science, pp. 333-351.

Anderson, T. W. (1984) *Introduction to multivariate statistical analysis*. New York: Wiley-Interscience.

Armah, F. A., Yawson, D. O. and Pappoe, A. A. (2010) ‘A systems dynamics approach to explore traffic congestion & air pollution link in the city of Accra, Ghana’, *Sustainability*, 2(1), 252-265.

Bader, D. A. and Cong, G. (2006) ‘Fast shared-memory algorithms for computing the minimum spanning forest of sparse graphs’, *Journal of Parallel and Distributed Computing*, 66(11), 1366-1378.

Bandelt, H. J., Forster, P. and Röhl, A. (1999) ‘Median-joining networks for inferring intraspecific phylogenies’, *Molecular Biology and Evolution*, 16(1), 37-48.

Bell, D. C., Atkinson, J. S. and Carlson, J. W. (1999) ‘Centrality measures for disease transmission networks’, *Social Networks*, 21(1), 1-21.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D. U. (2006) ‘Complex networks: Structure and dynamics’, *Physics Reports*, 424(4-5), 175-308.

Bonacich, P. (1972) ‘Factoring and weighting approaches to status scores and clique identification’, *Journal of Mathematical Sociology*, 2(1), 113–120.

Bondy, J. A. and ZMurty, U. S. R. (1976). *Graph theory with applications*. London: Macmillan.

Borgatti, S. P. (1995) ‘Centrality and AIDS’, *Connection*, 18(1), 112-115.

Borgatti, S. P. (2005) ‘Centrality and network flow’, *Social Networks*, 27(1), 55-71.

Borgatti, S. P., Everett, M. G. and Johnson, J. C. (2013) *Analyzing Social Network*. SAGE Publications.

Bounova, G. A (2009) Topological Evolution of Networks: Case Studies in the US Airlines and Language Wikipedias. PhD Thesis, Massachusetts Institute of Technology.

Brandes, U. and Fleischer, D. (2005) Centrality Measures Based on Current Flow. *Annual Symposium on Theoretical Aspects of Computer Science*. Springer Berlin Heidelberg. 533-544.

Brida, J. G. and Risso, W. A. (2010) 'Hierarchical structure of the German stock market', *Expert Systems with Applications*, 37(5), 3846-3852.

Broutin, N., Devroye, L. and McLeish, E. (2010) 'Note on the structure of Kruskal's algorithm', *Journal: Algorithmica*, 56(2), 141-159.

Chang, C. L. and Chang, S. K. (1977) Storage Reduction through Minimal Spanning Trees and Spanning Forests. *Proceeding of the 1977 IEEE Transactions on Computers*. May 1977. IEEE, 425-434.

Chatterjee, S. and Hadi, A.S. (1998) *Sensitivity analysis in linear regression*. New York: John Wiley & Sons.

Chen, D., Lü, L., Shang, M. S., Zhang, Y. C. and Zhou, T. (2012) 'Identifying influential nodes in complex networks', *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1777-1787.

Ching, N. H., Rosenfeld, D. and Braun, M. (1992) Two-Dimensional Phase Unwrapping using a Minimum Spanning Tree Algorithm. *Proceeding of the IEEE Transactions on Image Processing*. IEEE, 355-365.

Christofides, N., Mingozzi, A. and Toth, P. (1981) 'Exact algorithms for the vehicle routing problem, based on spanning tree and shortest path relaxations', *Mathematical Programming*, 20(1), 255-282.

Crucitti, P., Latora, V. and Porta, S. (2006) 'Centrality measures in spatial networks of urban streets', *Physical Review E*, 73(3), 036125.

Cupal, M., Deev, O. and Linnertová, D. (2012) Network Structures of the European Stock Markets. *Proceedings of 30th International Conference Mathematical Methods in Economics*. 79-84.

De Montis, A., Barth´elemy, M., Chessa, A. and Vespignani, A. (2007) 'The structure of inter-urban traffic: A weighted network analysis', *Environment and Planning B: Planning and Design*, 34, 905–927.

Dementiev, R., Sanders, P., Schultes, D. and Sibeyn, J. (2004) Engineering an External Memory Minimum Spanning Tree Algorithm. *Proceedings of the 2004 Exploring New Frontiers of Theoretical Informatics.* 195-208. doi: 10.1007/1-4020-8141-3_17.

Department of Statistics, Malaysia. (2016) *Statistical Handbook Malaysia*.

Djauhari, M.A. (2005) ‘Improved monitoring of multivariate process variability’, *Journal Quality of Technology*, 37(1), 32-39.

Djauhari, M. A. (2007) ‘A Measure of multivariate data concentration’, *Journal of Applied Probability and Statistics,* 2(2), 139-155.

Djauhari, M. A. (2012) ‘A robust filter in stock networks analysis’, *Physica A: Statistical Mechanics and Its Applications,* 391(20), 5049-5057. https://doi.org/10.1016/j.physa.2012.05.060.

Djauhari, M. A. (2012) Influential data subsets labelling: A fuzzy relation approach. *Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, Langkawi, 2012, 1-3. doi: 10.1109/ICSSBE.2012.6396633

Djauhari, M. A. and Gan, S. L. (2013) ‘Minimal spanning tree problem in stock networks analysis: An efficient algorithm’, *Physica A: Statistical Mechanics and Its Applications*, 392(9), 2226-2234. https://doi.org/10.1016/j.physa.2012.12.032.

Djauhari, M. A. and Gan, S. L. (2015) ‘Optimality problem of network topology in stocks market analysis’, *Physica A: Statistical Mechanics and Its Applications,* 419, 108-114. https://doi.org/10.1016/j.physa.2014.09.060.

Djauhari, M. A. and Lee, G. S. (2013) ‘Network topological property of English dialects similarity: A robust filter approach’, *American Journal of Applied Sciences*, 10(7), 646-653. doi:10.3844/ajassp.2013.646.653.

Djauhari, M. A., Mashuri, M. and Herwindiati, D. E. (2008) ‘Multivariate process variability monitoring’, *Communications in Statistics – Theory and Methods,* 37, 1742-1754. http://dx.doi.org/10.1080/03610920701826286.

Djauhari, M. A. and Mohamad, I. (2010) ‘How to control process variability more effectively: The case in B-complex vitamin production process’, *South African Journal of Industrial Engineering*, 21(21), 207-215.

Djauhari, M. A., Sharif, S. and Djauhari, H. (2012) ‘Network analysis on safety culture and worker‘s behaviour: A forest of all minimum spanning trees’, *International Journal of Basic and Applied Sciences,* 12(6), 29-37.

Djauhari, M. A. (2016) Advanced monitoring techniques for complex process variability in manufacturing industry. Universiti Putra Malaysia Press.

Djikstra E. W. (1959) ‘A note on two problems in connexion with graphs’, *Numerische Mathematik*, 1(1), 269-271. doi:10.1007/BF01386390.

Edelman, A. and Rao, N. R. (2005) ‘Random matrix theory’, *Acta Numerica*, 1-65.

Eom, C., Oh, G. and Kim, S. (2006) ‘Topological properties of the minimal spanning tree in Korean and American stock markets’, *Data Analysis, Statistics and Probability,* doi:10.3938/jkps.51.1432.

Eom, C., Oh, G., Jung, W. S., Jeong, H. and Kim, S. (2009) ‘Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series’, *Physica A: Statistical Mechanics and Its Application*, 388 (6), 900-906. https://doi.org/10.1016/j.physa.2008.12.006.

Everett, M. G. and Borgatti, S. P. (2013). ‘The dual-projection approach for two-mode networks’, *Social Networks*, *35*(2), 204-210.

Freeman, L.C. (1978) ‘Centrality in social networks: Conceptual clarification’, *Social Networks,* 1, 215–239. https://doi.org/10.1016/0378-8733(78)90021-7.

Gábor, S., and Csabai, I. (2002). ‘The analogies of highway and computer network traffic’, *Physica A: Statistical Mechanics and its Applications*, 307(3-4), 516-526.

Gan, S. L. and Djauhari, M. A. (2015) ‘New York stock exchange performance: Evidence from the forest of multidimensional minimum spanning trees’, *Journal of Statistical Mechanics: Theory and Experiment,* (12), P12005. doi:10.1088/1742-5468/2015/12/P12005.

Graham, R. L. and Hell, P. (1985) ‘On the history of the minimum spanning tree problem’, *Annals of the History of Computing*, 7(1), 43-57. doi: 10.1109/ MAHC.1985.10011.

Haibin, C. H. E. N. (2016) ‘Complex network analysis of highway network in Guangdong province in China’, *Canadian Social Science, 12*(3), 1-4.

Hanneman, R., & Riddle, M. (2005) *Introduction to social network methods*. E-book library [online]. Available at: http://faculty.ucr.edu/~hanneman/ (Accessed: 10 January 2014)

Hanslik, T., Boelle, P. and Flahault, A. J. (2001) ‘The control chart: An epidemiological tool for public health monitoring’, *Public Health*, 115, 277-281.

Heiberger, R. H. (2014) ‘Stock network stability in times of crisis’, *Physica A: Statistical Mechanics and Its Applications,* 393, 376-381. https://doi.org/10.1016/j.physa.2013.08.053.

Held, M. and Karp, R. M. (1970) ‘The traveling-salesman problem and minimum spanning trees’, *Operations Research*, 18(6), 1138-1162. https://doi.org/10.1287/opre.18.6.1138.

Held, M. and Karp, R. M. (1971) ‘The traveling-salesman problem and minimum spanning trees: Part II’, *Mathematical Programming*, 1(1), 6-25. doi:10.1007/BF01584070.

Herwindiati, D. E., Djauhari, M. A. and Mashuri, M. (2007) ‘Robust Multivariate Outlier Labeling’, *Communications in Statistics – Simulation and Computation*, 36(6), 1287-1294.

Huang, F., Gao, P. and Wang, Y. (2009) Comparison of Prim and Kruskal on Shanghai and Shenzhen 300 Index Hierarchical Structure Tree. *Proceedings of the 2009 International Conference on Web Information Systems and Mining*. November 7-8, 2009. Shanghai, China: IEEE, 237-241. doi: 10.1109/WISM.2009.56.

Hubert, M., Rousseeuw, P.J. and Van Aelst, S. (2008) ‘High-breakdown robust multivariate methods’, *Statistical Science*, 23(1). 92-119.

Jung, W. S., Kwon, O., Yang, J. S. and Moon, H. T. (2006) ‘Effects of globalization in the Korean financial market’, *Journal of the Korean Physical Society*, 48, 135-138. doi:10.1007/978-0-230-50575-9_9.

Kazemilari, M. and Djauhari, M. A. (2015) ‘Correlation network analysis for multi-dimensional data in stocks market’, *Physica A: Statistical Mechanics and Its Applications*, 429, 62-75. https://doi.org/10.1016/j.physa.2015.02.052.

Keskin, M., Deviren, B. and Kocakaplan, Y. (2011) ‘Topology of the correlation networks among major currencies using hierarchical structure methods’, *Physica A: Statistical Mechanics and Its Applications*, 390(4), 719-730. https://doi.org/10.1016/j.physa.2010.10.041.

Kruskal, J. B. (1956) On The Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*. 7(1), 48-50. doi: 10.2307/2033241.

Kwapień, J., Gworek, S., Drożdż, S. and Górski, A. (2009) ‘Analysis of a network structure of the foreign currency exchange market’, *Journal of Economic Interaction and Coordination*, 4(1), 55-72. doi:10.1007/s11403-009-0047-9.

Latora, V. and M. Marchiori. (2007) ‘A measure of centrality based on the network efficiency’, *New Journal of Physics*, 9(6), 188. doi:10.1088/1367-2630/9/6/188.

Lee, G. S., & Djauhari, M. A. (2012a) Network Topology of Indonesian Stock Market. *Proceedings of the 2012 IEEE on Cloud Computing and Social Networking*. April 26-27, 2012. Bandung, West Java, Indonesia: IEEE, 1-4. doi: 10.1109/ICCCSN.2012.6215721.

Lee, G. S. and Djauhari, M. A. (2012b) 'An overall centrality measure: The case of US stock market', *International Journal of Electrical & Computer Sciences,* 12(6). 99-103.

Lohmann, G., Daniel, S. M., Annette, H., Burkhard, P., Joeran, L., Dirk, G., Haiko, S., Michael, S., Arno, V. and Robert, T. (2010) 'Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain', *PloS one,* 5(4), e10232. https://doi.org/10.1371/journal.pone.0010232.

Lusiantoro, L. and Ciptono, W. S. (2012) 'An alternative to optimize the Indonesian's airport network design: An application of minimum spanning tree (MST) technique', *Gadjah Mada International Journal of Business*, 14(3), 267-302. 10.22146/gamaijb.5477.

Mantegna, R.N. (1999) 'Hierarchical structure in financial markets', *The European Physical Journal B,* 11(1), 193-197. doi:10.1007/s100510050929.

Mantegna, R.N. and Stanley, H.E. (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, United Kingdom.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*, Academic Press, New York.

Mason, R. L., Chou, Y. M. and Young, J. C. (2009) 'Monitoring variation in a multivariate process when the dimension is large relative to the sample size', *Communications in Statistics - Theory and Methods,* 38, 939-951. http://dx.doi.org/10.1080/03610920802339825.

Messmer, A., Papageorgiou, M. and Mackenzie, N. (1998) 'Automatic control of variable message signs in the interurban Scottish highway network', *Transportation Research Part C*, 6, 173-187. https://doi.org/10.1016/S0968-090X(98)00014-X.

Montgomery D. C. (2001) *Introduction to Statistical Quality Control*. (4^{th} ed.) New York: John Wiley & Sons, Inc.

Montgomery D. C. (2005) *Introduction to Statistical Quality Control*. (5^{th} ed.) New York: John Wiley & Sons, Inc.

Montgomery, D. C. (2009) *Introduction to Statistical Quality Control*. (6th ed.) New York: John Wiley & Sons, Inc.

Naylor, M. J., Rose, L. and Moyle, B.J. (2007) 'Topology of foreign exchange markets using hierarchical structure methods', *Physica A: Statistical Mechanics and its Applications*, 382(1), 199-208.

Nešetril, J. and Nešetrilová, H. (2010) 'The origins of minimal spanning tree algorithms–Boruvka and Jarnık', *Documenta Mathematica*, 127-141.

Newman, M. E. (2008) 'The mathematics of networks', *The New Palgrave Encyclopedia of Economics*, 2, 1-12.

Nobi, A., Maeng, S. E., Ha, G. G. and Lee, J. W. (2015) 'Structural changes in the minimal spanning tree and the hierarchical network in the Korean Stock market around the global financial crisis', *Journal of the Korean Physical Society*, 66(8), 1153-1159. doi:10.3938/jkps.66.1153.

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) 'Node centrality in weighted networks: Generalizing degree and shortest paths', *Social Networks*, 32(3), 245-251. https://doi.org/10.1016/j.socnet.2010.03.006.

Park, K. and Yilmaz, A. (2010) A Social Network Analysis Approach to Analyze Road Networks. *Proceedings of the 2010 ASPRS Annual Conference.* April 26-30, 2010. San Diego, CA. 1-6.

Pena, D. and Rodriguez, J. (2003) 'Descriptive measures of multivariate scatter and linear dependence', *Journal of Multivariate Analysis*, 85(2), 361–374. https://doi.org/10.1016/S0047-259X(02)00061-1.

Perra, N. and Fortunato, S. (2008) 'Spectral centrality measures in complex networks', *Physical Review E*, 78(3), 036107. https://doi.org/10.1103/PhysRevE. 78.036107.

Prim, R. C. (1957) 'Shortest connection networks and some generalizations', *Bell Labs Technical Journal*, 36(6), 1389-1401. doi:10.1002/j.1538-7305.1957.tb0151 5.x.

Reis, M. S. and Rato, T. J. (2013) Large-scale statistical process control: Should we use partial or marginal correlations? Proceeding of the American Institute of Chemical Engineers (AIChE). November 3- 8, 2013. Hotel San Francisco Union Square, San Francisco. https://www.aiche.org/resources/proceedings/ aiche-annual-meeting/2013.

Roy, R. B. and Sarkar, U. K. (2011) A Social Network Approach to Examine the Role of Influential Stocks in Shaping Interdependence Structure in Global Stock Markets. *Proceedings of the 2011 IEEE International Conference on Advances in Social Networks Analysis and Mining*. July 25-27, 2011. Kaohsiung, Taiwan: IEEE, 567-569. doi: 10.1109/ASONAM.2011.87.

Sandoval L. J. (2012) 'A map of the Brazilian stock market', *Advances in Complex Systems*, 15(5), 1–30. https://doi.org/10.1142/S0219525912500427.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons, Inc.

Shirley, K. and Bradley, K. D. (2009) 'Examining graduate committee faculty compositions - A social network analysis example'.http://www.uky.edu/~kdbrad2/Kate.pdf

Singh, A. (2009) 'An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem', *Applied Soft Computing*, 9(2), 625-631. https://doi.org/10.1016/j.asoc.2008.09.001.

Situngkir, H. and Surya, Y. (2005) 'Indonesian stock market dynamics through ultrametricity of minimum spanning tree', 1-18. http://dx.doi.org/10.2139/ssrn.768204.

Sohn, J. (2006) 'Evaluating the significance of highway network links under the flood damage: An accessibility approach', *Transportation Research Part A*, 40(6), 491–506. https://doi.org/10.1016/j.tra.2005.08.006.

Stam, C. J., Tewarie, P., Van Dellen, E., Van Straaten, E. C. W., Hillebrand, A. and Van Mieghem, P. (2014) 'The trees and the forest: characterization of complex brain networks with minimum spanning trees', *International Journal of Psychophysiology*. 92(3), 129-138. https://doi.org/10.1016/j.ijpsycho.2014.04.001.

Tabak, B. M., Serra, T. R. and Cajueiro, D. O. (2010) 'Topological properties of stock market networks: The case of Brazil', *Physica A: Statistical Mechanics and Its Applications*, 389(16), 3240-3249. https://doi.org/10.1016/j.physa.2010.04.002.

Tang, P.F. and Barnett, N.S. (1996) 'Dispersion control for multivariate processes', *Australian Journal of Statistics*, 38(3), 235-251.

Tapia, E. and Rojas, R. (2003). Recognition of On-Line Handwritten Mathematical Expressions Using a Minimum Spanning Tree Construction and Symbol

Dominance. *International Workshop on Graphics Recognition*. 329-340. doi: 10.1007/ 978-3-540-25977-0_30.

Tarabalka, Y., Chanussot, J. and Benediktsson, J. A. (2009) Classification of Hyperspectral Images Using Automatic Marker Selection & Minimum Spanning Forest. *Proceedings of the 2009 Hyper spectral Image and Signal Processing: Evolution in Remote Sensing.* August 26-28, 2009. Grenoble, France: IEEE, 1-4. doi: 10.1109/WHISPERS.2009.5289054.

Tarabalka, Y., Chanussot, J. and Benediktsson, J. A. (2009) Segmentation & Classification of Hyperspectral Images Using Minimum Spanning Forest Grown From Automatically Selected Markers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).* 40(5), 1267-1279. doi: 10.1109/ TSMCB.2009.2037132.

Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. (2005) A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences,* 102(30), 10421-10426.

Ministry of Transport (MOT). *Transport Statistics Malaysia*. (2016).

Ulusoy, T., Keskin, M., Shirvani, A., Deviren, B., Kantar, E. and Dönmez, C. Ç. (2012) ‘Complexity of major UK companies between 2006 and 2010: Hierarchical structure method approach’, *Physica A: Statistical Mechanics and Its Applications,* 391(21), 5121-5131. https://doi.org/10.1016/j.physa.2012. 01.026

Valente, T. W., Coronges, K., Lakon, C. and Costenbader, E. (2008) ‘How correlated are network centrality measures?’, *Connections (Toronto, Ont)*, 28(1), 16-26.

Villas, B.P.R., Rodrigues F.A. and Fontoura, C. L. (2009) *modeling highway networks with path-geographical transformations,* in: Fortunato S., Mangioni G., Menezes R., Nicosia V. (eds.) *Complex Networks. Studies in Computational Intelligence,* 207. Springer, Berlin, Heidelberg.

Wiliński, M., Sienkiewicz, A., Gubiec, T., Kutner, R. and Struzik, Z.R. (2013) ‘Structural and topological phase transitions on the German stock exchange’, *Physica A,* 392(23), 5963–5973. https://doi.org/10.1016/j.physa.2013.07.064.

Woodall, W. H. and Montgomery, D. C. (1999) ‘Research issues and ideas in statistical process control’, *Journal of Quality Technology,* 38, 89-104.

Wright, M., Horowitz, R. and Kurzhanskiy, A. A. (2016) A dynamic system characterization of road network node models. *IFAC-PapersOnLine*. 49(18), 1054-1059. https://doi.org/10.1016/j.ifacol.2016.10.307.

Xu, Y., Olman, V. and Xu, D. (2001) 'Minimum spanning trees for gene expression data clustering', *Genome Informatics*, 12, 24-33. http://doi.org/10.11234/gi 1990.12.24.

Xu, Y., Olman, V. and Xu, D. (2002) 'Clustering gene expression data using a graph-theoretic approach: An application of minimum spanning trees', *Bioinformatics*, 18(4), 536-545. doi: https://doi.org/10.1093/bioinformatics/ 18.4.536.

Yuewu, Y. (2010) A New Algorithm for Vehicle Routing Problems with Capacity Limited Based on Minimum Spanning Tree. *Proceedings of the 2010 IEEE Transactions on Optoelectronics and Image Processing.* November 11-12, 2010. Haiko, Hainan, China, China: IEEE, 47-50. doi: 10.1109/ICOIP.2010. 323.

Yusoff, N. S., Djauhari, M. A., Sharif, S. and Suleiman, E. S. (2012) 'Organizational commitment in Malaysian public university: An evidence via social network analysis, *International Journal of Basic and Applied Sciences*, 12(5), 17-21.

Zhang, Y., Lee, G. H. T., Wong, J. C., Kok, J. L., Prusty, M. and Cheong, S. A. (2011) 'Will the US economy recover in 2010? A minimal spanning tree study', *Physica A: Statistical Mechanics and Its Applications*, 390(11), 2020-2050. https://doi.org/10.1016/j.physa.2011.01.020.

LIST OF PUBLICATIONS

Indexed Journal

1. **Asrah, N. M.,** Djauhari, M. A., & Mohamad, I. (2016). PLUS Traffic Highway: An Analysis Based on Time Series Similarity Approach. The Social Sciences (Pakistan). 11(11), 2753-2759. doi:10.3923/sscience.2016.2753.2759. **(Indexed by SCOPUS)**

Indexed Conference Proceedings

1. **Asrah, N. M. &** Djauhari, M. A. A. (2014). Recommendation on PLUS Highway Development: A Social Network Analysis Approach. In *International Conference on Industrial Engineering and Engineering Management.* (pp.1187-1191). IEEE. doi:10.1109/IEEM.2014.7058826. **(Indexed by SCOPUS)**

2. **Asrah, N. M.,** Djauhari, M. A. & Mohamad, I. (2016). PLUS Highway Network Analysis: Case of In-Coming Traffic Burden in 2013. *In The 3rd ISM International Statistical Conference 2016.* (pp. 1-7). doi:10.1063/1.4982846 **(Indexed by SCOPUS)**

3. **Asrah, N. M., Djauhari,** M. A. & Mohamad, I. (2018). The Most Important Centrality Measures in PLUS Highway Traffic Network: An Approach Based on Effective Variance. In *Simposium Kebangsaan Sains Matematik ke 25 (SKSM25)*. doi:https://doi.org/10.1063/1.5041694 **(Indexed by SCOPUS)**

4. **Asrah, N. M., Djauhari,** M. A., Khamis, A., Salleh, R. M., & Avtar, S. S. S. M. F. (2018). Granger Causality Analysis on PLUS Highway Traffic Network. In *4th International Conference on The Application of Science and Mathematics 2018.* (pp. 432-437). https://www.sciencepubco.com/index.php/ijet/article/view/22359. **(Indexed by SCOPUS)**