Correlation Network Analysis on Worker's Behaviour and Safety Culture: An Experience in Manufacturing Industry

Shamshuritawati Sharif[1], Maman A. Djauhari[2] and Hariza Djauhari[3]
[1,2]Department of Mathematics, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia
[1]College of Arts and Sciences, UUM, 06010 UUM Sintok, Kedah, Malaysia
[3]Coordinating Ministry for Economic Affairs, Indonesia
[1]shamshurita@uum.edu.my, [2]maman@utm.my, [1]hariza.djauhari@yahoo.com

***Abstract*— A manufacturing industry contributes around 10% Malaysian economy. It provides economic opportunities for related industries and business. However, the number of accidents in manufacturing sector, including fatal accidents, has been increased from time to time. To understand the real situation, in this paper, we used a correlation network analysis to analyze 43 characteristics of worker's behaviour and their safety culture. The method developed in econophysics has been used to transform the correlation structure into distance structure. Its corresponding minimum spanning tree and the centrality measure such as degree centrality, betweenness centrality, closeness centrality and eigenvector centrality are performed to identify the most influential characteristics. A case study on Malaysian manufacturing industry has been presented to illustrate the advantage of the proposed approach. Some of important results and recommendations for the Government of Malaysia will be delivered.**

***Index Term*— adjacency matrix, complex system, correlation matrix, distance matrix.**

I. Introduction

Department of Occupational Safety and Health (DOSH) of Malaysia has to provide a safety and healthy work environment for all employees and protect those who may be affected by industry activities. DOSH activities are to guarantee employers and employees in the country pay more attention to safety and health at work [1].

On the other hand, however, the number of accidents in manufacturing sector, including fatal accidents, has been increased from time to time. In the last three years, Malaysia's manufacturing sector has been contributing the highest number of accidents which result in non-permanent disabilities (NPD), permanent disabilities (PD) and death (D). This sector becomes the second sector where accident occurrences causing death is placed on the top behind the construction sector as can be seen in Table 1. The occurrences of occupational accidents itself are believed due to direct, indirect and basic causes. Human factor is believed as the most contributors to the accident occurrences. In this regards [2] and [3] have mentioned that unsafe behaviour of the workers is the direct source of accident along with unsafe condition.

TABLE 1
The occupational accident by sector in 2008 – 2010

Sector	2008			2009			2010		
	NPD	PD	D	NPD	PD	D	NPD	PD	D
Manufacturing	1565	134	76	1186	79	53	1367	154	57
Agriculture, Forestry, Logging & Fishing	365	7	42	363	8	40	433	17	29
Utility	82	12	19	85	3	18	34	3	11
Construction	55	2	72	34	6	62	46	4	63
Transport, Storage & Communication	18	1	8	18	0	8	13	1	11
Hotel & Restaurant	13	1	1	18	0	0	20	0	0
Public Services & Statutory Bodies	3	1	2	0	0	1	36	2	3
Mining & Quarrying	4	0	6	2	1	2	2	1	0
Wholesale & Retail Trade	2	0	0	0	0	0	0	0	0
Financial, Insurance, Real Estate & Business Services	2	1	4	0	0	1	26	1	1

This paper aims to have a better understanding to what extent the current practice of safety culture and worker's behaviour intolerably differs from DOSH strategy. The remainder of the paper is designed as follows. Section II will be devoted to research design and implementation on safety culture and worker's behaviour followed by data analysis methodology in Section III. Later, in Section IV we discuss the research results. A conclusion will be delivered in the last section.

II. Research Methodology

In this study, safety culture and worker's behaviour are considered as a complex system consisting of 43 characteristics as nodes connected by (43-1)*43/2 = 903 links each of which is related to the correlation coefficient between the two nodes adjacent to it. The nodes and links constitute a social network in the form of a weighted undirected graph [4]. This point of view is useful in order to simplify, visualize, and summarize the most important information contained in that complex system.

In what follows we discuss first the research design, data collection, and data analysis based on the so-called minimum spanning tree (MST) and centrality measures.

A. Survey Design and Data Collection

There are 136 workers that have been participated in this survey. Our focus is on the front line workers only, i.e., operators and technicians because they are the main target of DOSH policy. The questionnaire consists of 43 questions (characteristics). Among them, the first 18 are related to safety culture and the rest to worker's behaviour.

The safety culture characteristics are classified into nine factors, namely; management commitment, communication, priority of safety, safety procedure and policy, supportive environment, involvement, personal priority and need of safety, personal appreciation towards risk, and work environment.

On the other hand, worker's behaviour is classified into seven factors; reacting behaviour, personal protective equipment, specific job risk, tools and equipments, safe work practice, ergonomics, and communication. See [5] for the details of the questionnaire.

B. Methodology

Those 43 nodes and 903 links can be considered as a social network, i.e., a network representation of social phenomenon viewed as a complex system. The essence of a network is its nodes and the way how they are linked. Network analysis was originally developed in computer science. However, nowadays, it has been used in various fields of study. See, for example, [6] in sociology, [7] and [8] in finance, and [9] in transportation.

In practice, network analysis might start with a correlation matrix. Then, we transform it into a distance matrix [10]. From that matrix we construct the minimum spanning tree (MST). For this purpose, as suggested in [7] and [10], Kruskal algorithm can be used. MST will then be used to filter the original network and summarize the most important information.

Furthermore, to interpret the MST we use dot plot matrix, and centrality measures such as degree, betweenness, closeness, and eigenvector centralities. To make the MST more attractively and efficiently useful, we use the Kamada Kawai procedure provided in Pajek [11].

III. Data Analysis

A. Correlation matrix

We denote X_i is the *i*-th characteristic under study where *i* = 1, 2, … , 43. The correlation matrix among those characteristics, issued from a sample, is a symmetric matrix of size 43×43 where the element in the *i*-th row and *j*-th column is,

$$\rho_{ij} = \frac{\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle}{\sqrt{\left(\langle X_i^2 \rangle - \langle X_i \rangle^2\right)\left(\langle X_j^2 \rangle - \langle X_j \rangle^2\right)}} \qquad (1)$$

representing the correlation coefficient between *i*-th and *j*-th characteristics [10]. That correlation coefficient quantifies the degree of linear relationship between *i*-th and *j*-th variables. By definition, $\rho_{ii} = 1$ for all *i* and ρ_{ij} can vary from – 1 to 1 for all $i \neq j$ where,

$$\rho_{ij} = \begin{cases} 1 & \text{means perfectly positive linear relationship} \\ 0 & \text{means no linear relationship} \\ -1 & \text{means perfectly negative linear relationship} \end{cases}$$

B. Distance matrix

To analyze the network, we transform the correlation matrix into a distance matrix by using the following formula [10].

$$d_{ij} = \sqrt{2(1-\rho_{ij})} \qquad (2)$$

This d_{ij} is a distance between the *i*-th and *j*-th characteristics since it satisfies the following three properties; (i) $d_{ij} \geq 0$ and $d_{ij} = 0 \Leftrightarrow X_i = X_j$, (ii) $d_{ij} = d_{ji}$, and (iii) $d_{ij} \leq d_{ik} + d_{kj}$. The first property tells us that two characteristics that are perfectly correlated (either positive or negative), $|\rho_{ij}| = 1$, will be represented by a single point in Euclidean space ($d_{ij} = 0$). More over, $0 \leq d_{ij} \leq 2$.

The second property is symmetric property; the distance between the *i*-th and *j*-th characteristics is equal to the distance between the *j*-th and *i*-th characteristics. In other words, the correlation between the *i*-th and *j*-th characteristics is equal to the correlation between the *j*-th and *i*-th characteristics ($\rho_{ij} = \rho_{ji} \Leftrightarrow d_{ij} = d_{ji}$).

The last property is well known as triangular property. From (2), we conclude that, in general, the higher the correlation coefficient the smaller the distance.

By using equation (2), we obtain a distance matrix D of size 43×43 with d_{ij} as the element in the *i*-th row and *j*-th column. It is this matrix that we analyze in the rest of the paper.

C. Information Summarization

To visualize, simplify and summarize the important information contained in the network represented by D, we use the notion MST [12] [13]. Then, we determine MST by using Kruskal algorithm [14].

Furthermore, to interpret the MST, dot plot matrix and centrality measures are used. These measures are very helpful to understand the importance and or influence of each node relative to the others [15], [16], and [17]. The role of each measure in details and its formula are discussed in [18].

IV. Result and Discussion

A. Minimum Spanning Tree

In Fig. 1 we present the dot plot matrix of the adjacency matrix A that corresponds to the MST of distance matrix D given by Kruskal's algorithm. The element of A is a_{ij} = 1 if the *i*-th and *j*-th nodes are linked and 0 otherwise. This matrix is symmetric and all diagonal elements are 0. In Fig. 1, blank cell represents 0 and black cell 1.

That figure shows the highest correlations among safety culture characteristics (the first eighteen rows and columns) are concentrated along diagonal while worker's behaviour characteristics (the remaining rows and columns) are more dispersed around diagonal. Moreover, some of worker's behaviour characteristics are also highly correlated with some safety culture characteristics. This indicates that managing worker's behaviour is more complicated compared to safety culture.

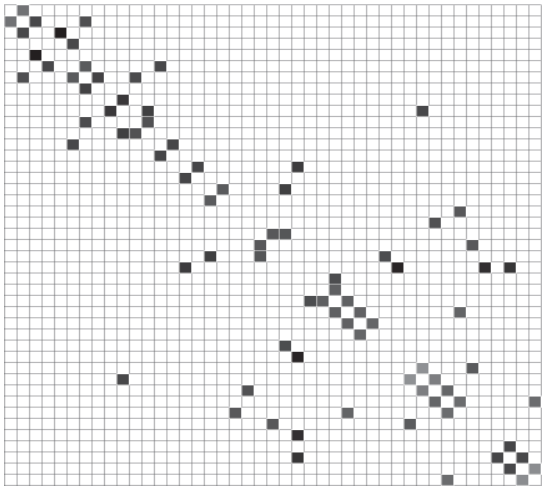

Fig. 1. Dot plot matrix

From that figure we also learn that:

(i) There is high correlation among characteristics within factors in safety culture but low correlation between factors.

(ii) There is high correlation among characteristics within and also between factors in worker's behaviour.

(iii) The following safety culture characteristics BE2, BH1, and BI1 (supportive environment, personal appreciation towards risk, and work environment factors) are highly correlated with the following worker's behaviour characteristics CD1, CC2, and CA3, (reacting behaviour, specific job risk, and tools and equipments factors), respectively.

To elaborate the above findings more clearly, we use Pajek software to represent in Fig. 2 the corresponding MST. See, [19] for consulting the open source. From this figure we see the interconnectivity among all characteristics of both groups. Yellow points represent safety culture characteristics and red points are for worker's behaviour characteristics. In general, the two groups are clearly separated except the following two factors of safety culture that are more similar with factors of worker's behaviour than within its own group,

(i) personal appreciation towards risk (BH1 and BH2),
(ii) work environment (BI1 and BI2).

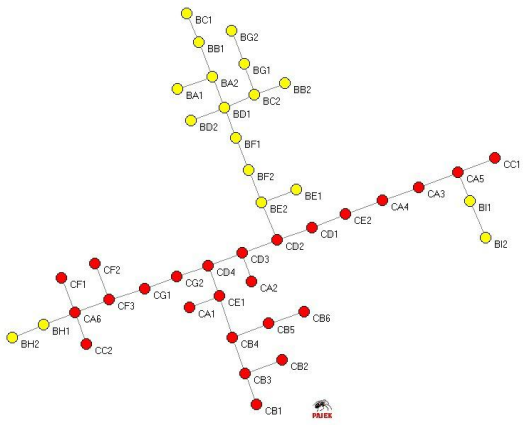

Fig. 2 MST of safety culture and worker's behaviour

Furthermore, according to Fig. 1, BE2 (supportive environment) in the yellow group is more similar to red group than yellow group. We conclude that Fig. 1 and Fig. 2 give us a general relationship among characteristics of the two groups.

In what follows we analyze the particularity of each characteristics by using the notion of centrality measure.

B. Centrality Measures

From social network view point, each particular node can be analyzed by using its centrality measures such as degree, betweenness, closeness and eigenvector centralities to find the most important nodes in the network structure. Those measures are computed based on the MST in Fig. 2. See [9], [18], and [20].

(i) Degree centrality of node i is $d_i = \frac{1}{n-1}\sum_{j=1}^{n} a_{ij}$.

(ii) Betweenness centrality of node i is b_i , the ratio of the number of path passing through i between two different nodes and the number of all possible paths from j to k for all j and k where $j \neq i$ and $k \neq i$.

(iii) Closeness centrality of node i, c_i is the ratio of the number of links in the MST (n-1) and the number of links in the path from i to j for all $j \neq i$.

(iv) Eigenvector centrality of node i is, $ev_i = \lambda^{-1}\sum_{j=1}^{n} a_{ij} e_j$ where $(e_1, e_2, ..., e_n)^t$ is the eigenvector of A that corresponds to the largest eigenvalue λ .

In Table 2, we present the value of those measures of each node.

TABLE 2
Centrality Measures

No.	i	d_i	b_i	c_i	ev_i
1	BA1	0.024	0	0.12	0.1
2	BA2	0.071	0.138	0.136	0.241
3	BB1	0.048	0.048	0.121	0.121
4	BB2	0.024	0	0.12	0.1
5	BC1	0.024	0	0.108	0.05
6	BC2	0.071	0.138	0.136	0.241
7	BD1	**0.095**	0.373	0.153	**0.358**
8	BD2	0.024	0	0.133	0.149
9	BE1	0.024	0	0.167	0.101
10	BE2	0.071	0.452	0.199	0.243
11	BF1	0.048	0.372	0.167	0.231
12	BF2	0.048	0.396	0.183	0.197
13	BG1	0.048	0.048	0.121	0.121
14	BG2	0.024	0	0.108	0.05
15	BH1	0.048	0.048	0.117	0.052
16	BH2	0.024	0	0.105	0.022
17	BI1	0.048	0.048	0.108	0.009
18	BI2	0.024	0	0.097	0.004
19	CA1	0.024	0	0.153	0.102
20	CA2	0.024	0	0.175	0.121
21	CA3	0.048	0.177	0.133	0.026
22	CA4	0.048	0.215	0.148	0.045
23	CA5	0.071	0.138	0.12	0.018

24	CA6	**0.095**	0.182	0.131	0.104
25	CB1	0.024	0	0.123	0.051
26	CB2	0.024	0	0.123	0.051
27	CB3	0.071	0.094	0.14	0.123
28	CB4	0.071	0.222	0.159	0.194
29	CB5	0.048	0.048	0.139	0.097
30	CB6	0.024	0	0.122	0.041
31	CC1	0.024	0	0.107	0.007
32	CC2	0.024	0	0.116	0.043
33	CD1	0.048	0.285	0.188	0.153
34	CD2	0.071	**0.641**	**0.214**	0.286
35	CD3	0.071	0.528	0.211	0.292
36	CD4	0.071	0.577	0.204	0.295
37	CE1	0.071	0.292	0.18	0.246
38	CE2	0.048	0.251	0.167	0.082
39	CF1	0.024	0	0.116	0.043
40	CF2	0.024	0	0.128	0.046
41	CF3	0.071	0.257	0.146	0.111
42	CG1	0.048	0.285	0.163	0.118
43	CG2	0.048	0.316	0.182	0.171

To visualize the result in Table 2, in Fig. 3 – 6 we present the MST where the size and colour of the node represent the score of centrality measure and the rank of importance, respectively.

Based on degree centrality, see Fig. 3, BD1-*Rules and Procedures on Health and Safety Programme* and CA6-*Seriousness in Work* (red points) have the highest number of links (4) in the network. Each of the followings has 3 links: BA2-*Management Responsiveness*, BC2-*Concern on Safety*, BE2-*Health and Safety Performance*, CA5-*Safe Environment*, CB3-*Hearing Protection*, CB4-*Respiratory Protection*, CD2-*Tools and Equipments Used Correctly*, CD3-*Tools and Equipment Well Maintain*, CD4-*Workplace in Good Housekeeping*, CE1-*Understand Work Safely*, and CF3-*Zero Awkward Posture* (yellow points).

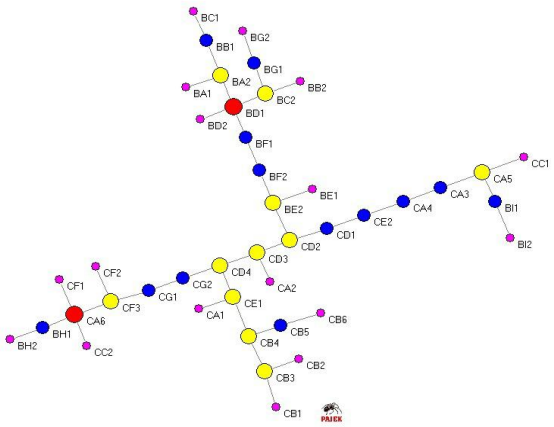

Fig. 3 Degree centrality

The rests are of 1 and 2 links only. The higher the number of links, the higher the influence of a particular characteristic. For example, BD1-*Rules and Procedures on Health and Safety Programme* is the most influential characteristic to the larger number of other characteristics of safety culture.

In terms of betweenness, see Fig. 4, the four most important nodes are CD2-*Tools and Equipments Used Correctly* (red point) plays the most important role in the network followed by, in order of importance: CD4-*Workplace in Good Housekeeping*, and CD3-*Tools and Equipment Well Maintain* (yellow points or the second most important), BE2-*Health and Safety Performance* (blue point or the third most important), BF2-*Participate in Work Safety Issues*, BD1-*Rules and Procedures on Health and Safety Programme*, BF1-*Participate in Management Safety Issues*, CG2-*Concern on Work Environment* (purple points or the fourth most important). This means that if those variables are well managed, then the others will be well influenced.

According to the closeness centrality, see Fig. 5, CD2-*Tools and Equipments Used Correctly* (red point) has an excellent position compared to the others where the information flow in the network can easily reach others. This node is the closest node to the others. The second (third, and fourth, respectively) closest node to the others are CD3-*Tools and Equipment Well Maintain* (yellow point) (CD4-*Workplace in Good Housekeeping* (blue point), and all purple points, respectively).

In term of eigenvector centrality, see Fig. 6, BD1-*Rules and Procedures on Health and Safety Programme* (red point) is the node that links to the other high scored nodes. It is the most influential characteristic to the second most influential characteristics CD2-*Tools and Equipments Used Correctly*, CD3-*Tools and Equipment Well Maintain*, and CD4-*Workplace in Good Housekeeping* (yellow points). The third most influential are the blue points.

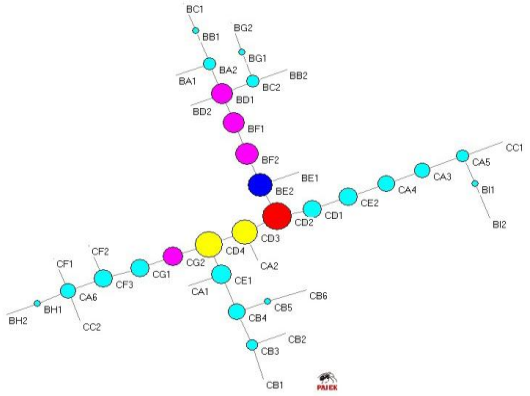

Fig. 4. Betweenness centrality

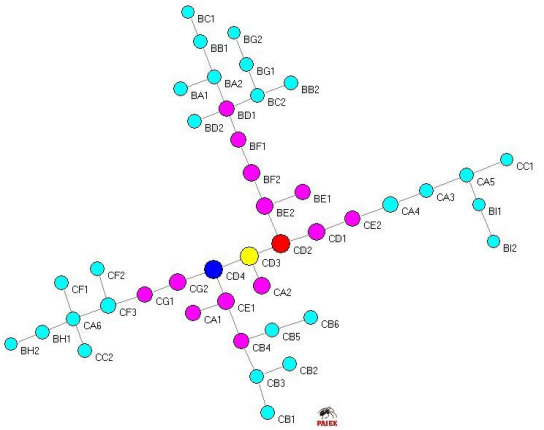

Fig. 5. Closeness centrality

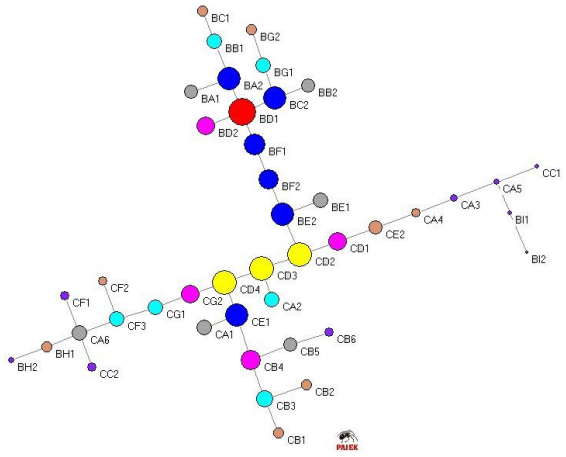

Fig. 6. Eigenvector centrality

V. Conclusion

Dot plot matrix analysis shows that managing worker's behaviour is more difficult to manage than safety culture because,

(i) the characteristics of the latter is highly correlated within factors but not between them, and
(ii) the characteristics of the former is highly correlated within and between factors and also with the factors of the former.

From Fig. 2 we learn that the factors of safety culture are clearly separated from those of worker behaviour except 'Personal appreciation toward risk' and 'Work environment' that are more similar to the second group of factors rather than their own group. According to the four centrality measures, after using the Pareto analysis, the following characteristics of safety culture are the vital few; BE2-*Health and Safety Performance*, BD1-*Rules and Procedures on Health and Safety Programme*, BF2- *Participate in Work Safety Issues*, BA2-*Management Responsiveness*, and BC2-*Concern on Safety*. These characteristics represent the following factors in order of importance,

(i) supportive environment
(ii) safety procedure and policy
(iii) involvement
(iv) management commitment
(v) priority of safety

On the other hand, the vital few of worker's behaviour characteristics are CD2-*Tools And Equipments Used Correctly*, CD3-*Tools and Equipment Well Maintain*, CD4-*Workplace in Good Housekeeping*, CE1-*Understand Work Safely*, CD1-*Right Tools and Equipments*, CG2-*Concern on Work Environment*, CA2- *Follow a Regulation* that cover the factors,

(i) tools and equipment
(ii) safe work practise
(iii) communication
(iv) reacting behaviour

Those five factors of safety culture as well as those four factors of worker's behaviour should be of high priority in reducing the number of fatality and paid more attention by DOSH as well as Malaysian industrial management

I. Concluding Remarks

In this paper, we show that correlation network analysis can be used in identifying the most influential variables, in the case of several interrelated variables. For that purpose, it is not an easy task when involved large data set with high dimension. To close this paper, we highlight three potential topics from this study. First, it is still in our investigation to define the new measure of centrality due to some limitation of degree centrality. Second, it will be more fruitful to explore the limitation of the algorithm (i.e Kruskal algorithm) in order to determine a minimum spanning tree since that algorithm will only give us a unique solution. Third, we can also use the notion of MST in order to see the separation of suspected outliers as well as outlier testing. More studies can be carried out to investigate the limitation of this topic.

Acknowledgment

The authors gratefully acknowledge Goverment of Malaysia for those sponsorships, Universiti Teknologi Malaysia and Universiti Utara Malaysia for the facilities. Special thanks go to the anonymous referees for their constructive comments and suggestions.

References

[1] (2010) Malaysia Economic Report. [Online]. Available: http://www.treasury.gov.my.
[2] D. Heberle, *Construction Safety Manual.* New York: McGraw-Hill Professional, 1998.
[3] J. Short, The role of safety culture in preventing commercial motor vehicle crashes. *Volume 14 of Synthesis (Commercial Truck and Bus Safety Synthesis Program (U.S.))*. Washington D. C. Transportation Research Board, 2007.
[4] P. Jayawant and K. Glavin, "Minimum spanning trees," *Involve a journal of mathematics*, vol.2, no.4, pp. 439–450, 2009.
[5] H. Djauhari, "Safety Culture at a manufacturing Industry in Johor Malaysia," *Master Dissertation*. Universiti Teknologi Malaysia, Johor, Malaysia, 2010.
[6] T. Krichel, and N. Bakkalbasi, "A Social Network Analysis of Research Collaboration in the Economics Community," *The International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*, France, 2006.
[7] R. N. Mantegna, "Hierarchical Structure in Financial Markets," *European Physical Journal B*, vol. 11, pp. 193-197, 1999.
[8] S. Micciché, G. Bonanno, F. Lillo, and R.N. Mantegna, "Degree stability of a minimum spanning tree of price return and volatility," *Physica A*, vol. 324, pp.66–73, 2003.
[9] K. Park, and A. Yilmaz, "A Social Network Analysis Approach to Analyze Road Networks," *ASPRS Annual Conference*. San Diego, CA, 2010.
[10] R. N. Mantegna, and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge UK, 2000.
[11] T. Kamada, S. Kawai, "An algorithm for drawing general undirected graphs". *Information Processing Letters*, vol. 31, no.1, pp. 7–15, 1989.
[12] J. P. Benzécri *Histoire et préhistoire de l'Analyse des Données*, DUNOD: Paris, 1982.
[13] B. M. Tabak, T. R. Serra, and D. O. Cajueiro. Topological properties of stock market networks: The case of Brazil. *Physica A*, vol. 389, pp. 3240-3249, 2010.

[14] J.B. Kruskal, *"On the shortest spanning subtree and the travelling salesman problem". Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48-50, 1956.
[15] Y. Xu, J. Ma, Y. Sun, J. Hao, Sun, Y. And Y. Zhao, "Using Social Network Analysis As A Strategy For E-Commerce Recommendation," *Pacific Asia Conference on Information Systems (PACIS)*. India, 2009.
[16] A. Abbasi, and J. Altmann, "On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks". *TEMEP Discussion Paper, No.2010*, Seoul National University, Korea, 2010.
[17] J. Monárrez-Espino, and J. R. Caballero-Hoyos. "Stability of Centrality Measures in Social Network Analyses to Identify Long-Lasting Leaders from an Indigenous Boarding School of Northern Mexico," *Estudios sobre las Culturas Contemporaneas*, vol. 16, no. 32, pp.155-171, 2010.
[18] S. P. Borgatti, "Centrality and AIDS," *Connections*, vol. 18, no. 1, pp.112-114, 1995.
[19] V. Batagelj, and A. Mrvar, (2011) PAJEK: Program for Analysis and Visualization of Large Networks, version 2.02. [online]. Available: http://pajek.imfm.si/doku.php?id=download.
[20] P. Sieczka, and J. A. Holyst, "Correlations in commodity markets," *Physica A*, vol. 388, pp.1621-1630, 2009.

Shamshuritawati Sharif was born in Kedah, Malaysia, in 1977. She is currently a PhD student in Mathematics at Universiti Teknologi Malaysia, Malaysia. Her diploma and Bachelor of Science in Statistics were obtained from MARA Institute of Technology (ITM). Her Master degrees in Decision Science were from Universiti Utara Malaysia in 2003. She is interested in multivariate analysis, industrial statistics, and social network analysis.