Correlation Network Analysis of International Postgraduate Students’ Satisfaction in Top Malaysian Universities: A Robust Approach

Shamshuritawati Sharif1,2, Ahmadreza Shekarchizadeh3, Maman Abdurachman Djauhari1 & Amran Rasli4

1 Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Malaysia
2 College Arts and Sciences, Universiti Utara Malaysia, Sintok, Kedah, Malaysia
3 Department of Management, NajafAbad Branch, Islamic Azad University, Isfahan, Iran
4 Faculty of Management and HRD, Universiti Teknologi Malaysia, Malaysia

Correspondence: Shamshuritawati Sharif, Faculty of Science, Universiti Teknologi Malaysia, Skudai 81100 UTM, Johor, Malaysia. Tel: 601-9-424-8001. E-mail: shamshurita@uum.edu.my

Received: September 21, 2012     Accepted: November 22, 2012     Online Published: November 28, 2012
doi:10.5539/mas.v6n12p91          URL: http://dx.doi.org/10.5539/mas.v6n12p91

Abstract
This paper deals with an analysis of correlation network among the satisfaction characteristics evaluated by the international students of top five Malaysian universities. The method developed in econophysics has been used to filter the information contained in a correlation structure. In this paper, we present a correlation network analysis where the correlation structure is determined based on robust approach. This approach is used to have a better understanding of foreign students’ satisfaction which will be very useful for Malaysian universities. Based on our proposed centrality measure, interestingly, the result demonstrates that the item “pay premium price” is the most influential characteristic that reflects to the highest degree of satisfaction for the international students. The university senior management would benefit by knowing this result in order to gain competitive advantage over other universities. Practical implications are provided accordingly.

Keywords: adjacency matrix, Cronbach’s alpha, minimum spanning tree, outlier labelling

1. Introduction
Measuring consumer satisfaction in a competitive marketplace becomes an important differentiator of marketing strategy. Many consumer satisfaction definitions use a disconfirmation paradigm to measure satisfaction. By measuring consumer satisfaction, organizations are capable to get an evidence of how successful they actually are in providing services to the market.

Many customer satisfaction definitions use a disconfirmation paradigm to measure the antecedents of satisfaction. This paradigm states that satisfaction depends on the size and direction of the disconfirmation experience, whereas disconfirmation is a function of the person’s initial expectations. A person’s expectations are: (1) confirmed when a product performs as expected, (2) negatively disconfirmed when the product performs more poorly than expected, and (3) positively disconfirmed when the product performs better than expected (Churchill & Suprenant, 1982; Tse & Wilton, 1988).

Another follower of this school of thought was Oliver (1993), who maintains that customer satisfaction is related to prior expectations. In earlier works, Oliver (1989) identified several types of satisfaction: satisfaction-as-contentment, satisfaction as pleasure, satisfaction-as-relief, satisfaction-as-novelty and satisfaction-as-surprise. Later, the same author depicted another definition for satisfaction. He pointed out that satisfaction is the consumer’s fulfillment response. It is a judgment that a product or service feature, or the product or service itself, provided (or is providing) a pleasurable level of consumption-related fulfillment, including levels of under-or-over fulfillment (Oliver, 1997). In the context of the present paper, attitudes towards the quality of higher education are considered to affect student satisfaction.

According to Oliver (1981), customer satisfaction is relatively momentary and is consumption-specific, while attitudes are relatively enduring. In line with this thinking, Westbrook and Oliver (1981) argued that satisfaction is an assessment of the totality of the purchase situation compared to expectations, whereas an attitude is an
appreciating for a product or service that lacks the element of assessment. Therefore, it appears from this perspective that the level of satisfaction may vary depending on the alternatives available to customers.

For the purpose of the present research, student satisfaction is defined as an evaluating summary of educational experience, based on the discrepancy between prior expectation and the performance perceived after passing through the educational cycle.

Measuring student satisfaction is not an easy task to attempt. Therefore, the indicators that are used differ from one author to another. For example, Browne et al. (1998) found that global satisfaction within a university was driven by a student’s assessment of course quality and other curriculum-related factors associated with a university. Borden (1995) found that student satisfaction is related to the match between student priorities and the campus environment (Elliott & Shin, 2002). After reviewing the literature, it was revealed that student satisfaction causes three behavioral intentions: (1) repurchase, (2) word of mouth, and (3) pay price premium. Six items were developed based on these three behavioral intentions for measuring student satisfaction in a tertiary educational institution. A summary of the related characteristics and the researchers are presented in Table 1.

Table 1. Items for measuring international student’s satisfaction

<table>
<thead>
<tr>
<th>Code</th>
<th>Statement</th>
<th>Characteristics</th>
<th>Researcher(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Overall, I am pleased with my experience at my university</td>
<td>Good experience</td>
<td>Hampton, 1993</td>
</tr>
<tr>
<td>Q2</td>
<td>Overall, I am pleased with service quality at my university</td>
<td>Service quality</td>
<td>Harris, 2002</td>
</tr>
<tr>
<td>Q3</td>
<td>I would choose my university again for further studies (e.g. post doctoral)</td>
<td>Repurchase 1</td>
<td>Boulding, 1993</td>
</tr>
<tr>
<td>Q4</td>
<td>I would choose my university for another degree (for example after completing your PHD in engineering, you continue to do your MBA in the same university)</td>
<td>Repurchase 2</td>
<td>Boulding, 1993</td>
</tr>
<tr>
<td>Q5</td>
<td>I would recommend this university to a friend, relative, or potential student</td>
<td>Word mouth to mouth</td>
<td>Boulding, 1993; Ham, 2003</td>
</tr>
<tr>
<td>Q6</td>
<td>I would consider making a financial gift to my university</td>
<td>Pay premium price</td>
<td>Ham, 2003</td>
</tr>
</tbody>
</table>

The rest of the present paper is organized as follows. In the next section, we present the research methodology. Then, in Section 3, we will discuss the results of analysis. This paper will be closed with the conclusion in the last section.

2. Research Methodology

2.1 Data Preparation

There are 522 International postgraduate students involved in this study. The sample is collected based on stratified sampling technique at top five Malaysian public universities. The population for this study are students who enrolled for the second semester of 2008-2009. The top five universities are University Malaya (UM), University Teknologi Malaysia (UTM), Universiti Putra Malaysia (UPM), Universiti Kebangsaan Malaysia (UKM), and Universiti Sains Malaysia (USM). The measurement tool was based on questionnaire improved by Shekarchizadeh (2011). His questionnaire includes six behavioural intentions to assess students’ satisfaction. These items to measure students satisfaction are as presented in Table 1.

2.2 Robust Approach to Remove Outlier

A data which is inconsistent relative to the others in the group of data is called outlier (Barnet & Lewis, 1994). Identifying the outlier is very important issue in data analysis. The outliers are expected to exist in dataset with many observations and/or variables. The occurrences of the outliers can make the estimation of correlation matrix become inadequate.

In multivariate data setting, there are many procedures for identifying outliers. The most popular transformation is the Mahalanobis Squared Distance (MSD). A large value of MSD may indicate that corresponding observation is an outlier. However, Hadi (1992) mentioned that the outliers do not necessarily have large value of MSD and not all observations with large MSD value are necessarily outliers. These problems are known as masking and swamping effect due to the fact that mean vector and covariance matrix are not robust.
A classical method have disadvantage compared to robust estimation methods, which they are heavily influenced by outlier data points. Therefore, a robust approach is necessary to remove the outlier. In this paper, we implement the Minimum Covariance Determinant (MCD) estimator of Rousseeuw and Van Driessen (1999) for detecting and removing outlier. To see the advantage of MCD approach, we use a Cronbach’s alpha in order to measure the reliability and the internal consistency among all of six characteristics, following which, we can perform the network analysis to present the interrelationship among the characteristics of student’s satisfaction.

2.3 Correlation Network Analysis

Network analysis was originally developed in computer science. However, recently, it has been used in various fields of study. For instance, Krichel and Bakkalbasi (2006) in sociology, Mantegna (1999) and Miccichè et al. (2003) in finance, and Park and Yilmaz (2010) in transportation, utilized from this analysis.

In this research, the students’ satisfaction can be considered as a system consisting of 6 characteristics. If \( p \) represents number of nodes, then these nodes could be connected by \((p-1)p/2 = (6-1)6/2 = 15\) links. Each of these nodes is related to the correlation coefficient between the two nodes adjacent to it. The nodes and links will be considered as a network or, more specifically, a weighted undirected graph (Jayawant & Glavin, 2009). Those 6 nodes and 15 links can be considered as a social network.

Like in econophysics, network analysis was started in this study with a correlation matrix. Based on the recommendation by Mantegna (2000), the network analysis was transformed into distance matrix. From this matrix, the corresponding subdominant ultrametric (SDU) distance matrix was extracted based on Minimum Spanning Tree (MST). For this purpose, the Kruskal algorithm was used (Kruskal, 1956) as suggested in Mantegna (2000) and Jayawant and Glavin (2009). Then, MST was used to simplify the original network and summarize the most important information. To visualize the MST, an open source called ‘Pajek’ was used (Batagelj & Mrvar, 2003; De Nooy et al., 2004; Batagelj & Mrvar, 2011). In order to interpret the MST, centrality measures were utilized (Borgatti, 1995). Finally, the result of robust approach was compared with non robust approach. To make the MST more attractively and efficiently useful, the Kamada Kawai procedure provided in Pajek was utilized (Kamada & Kawai, 1989).

2.4 Centrality Measure

From network analysis view point, the role of importance of each particular node can be analyzed based on the four measures of centrality are commonly used in network analysis, i.e., degree, closeness, betweenness, and eigenvector centrality. These measures will help researcher to find the most important nodes in the network structure (Xu et al., 2009; Abbasi & Altmann, 2010; Monárrez-Espino & Caballero-Hoyos, 2010). Those measures are computed based on the MST as per the recommendations by Borgatti (1995), Sieczka and Holyst (2009) and Park and Yilmaz (2010).

There are many interesting research issues can be investigate in the area of centrality measure. For further discussion on centrality measure we can consult Coste badera and Valente (2003) who study the performance of centrality measure when networks are sampled, and Borgatti et al. (2006) who consider the robustness of basic network under the condition of imperfect data. However, our motivation goes to the new formulation of centrality measure since degree centrality is not sufficient to enrich the interpretation.

The degree centrality \( d_i \) was utilized to indicate the connectivity of nodes. It provides information on how many other nodes are connected with a particular node.

\[
d_i = \sum_{j=1}^{n} a_{ij}
\]

where \( a_{ij} = 1 \) if the \( i \)-th and \( j \)-th nodes are linked and 0 otherwise.

Degree centrality is the simplest of the node centrality measures by using the local structure around nodes only. For instance, as depicted in Figure 1, node \( d \) has 4 links and node \( e \) has 2 links.
In order to identify the role of importance, degree centrality is no longer appropriate to be the best measure. The higher the degree centrality does not reflect to the strength of each particular node. For improving the index for importance measurement, “average of weights” was proposed in this study. It is the average of weights $\bar{w}_j$ of all links adjacent to each node.

$$\bar{w}_j = \frac{1}{d_j} \sum_{i} w_{ij}$$

where $w_{ij}$ is weight of an $a_{ij}$.

This measure reflects the strength influence of a particular node to the others. The node that has larger scores in all measures is considered to be more central in terms of its influence to the others. For instance, nodes $d$ and $e$ have an average of weights 2 and 4.5, respectively (See Figure 2).

3. Results and Discussion

3.1 Correlation Network Analysis

The dataset of 522 international students has been screened for their validity, and only 487 respondents are selected to be used in the analysis. Subsequently, we used MCD which was formulated by Rousseeuw and Van Driessen (1999) to remove the outlier in order to avoid skewness and erroneous results. It was uncovered that about 119 outliers exist amongst the dataset. Based on the results of our research, we present the correlation analysis results in Table 2 by utilizing Matlab and LIBRA. Both of the results are representing robust and non robust correlation respectively.
Table 2. Correlation analysis

<table>
<thead>
<tr>
<th></th>
<th>(a): after correction</th>
<th></th>
<th>(b): before correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>Q1</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.695</td>
<td>Q2</td>
<td>0.653</td>
</tr>
<tr>
<td>Q3</td>
<td>0.581</td>
<td>Q3</td>
<td>0.510</td>
</tr>
<tr>
<td>Q4</td>
<td>0.556</td>
<td>Q4</td>
<td>0.480</td>
</tr>
<tr>
<td>Q5</td>
<td>0.493</td>
<td>Q5</td>
<td>0.434</td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td>Q6</td>
<td>0.434</td>
</tr>
</tbody>
</table>

The left section of Table 2 (after correction) shows that all the correlation coefficients have been changed after removing all the outliers as compared to the right section of Table 2 (before correction). As it can be observed, the removal of the outliers have caused the degree of relationship between Q1 and Q6 to increase, whilst, the degree of relationship between Q3 and Q4 has decreased. To further improve the analysis, Cronbach’s alpha was used for verification purposes. According to Cronbach and Shavelson (2004), the Cronbach’s alpha is a measure of reliability and internal consistency among all characteristics in our study. From the results it is apparent that the reliability coefficient has increased while using robust approach (alpha=0.917) as compared to the non robust approach (alpha=0.887). Since we expect the larger value of alpha, it shows that robust approach is applicable in this analysis.

Additionally, based on correlation matrix, we performed the network analysis to present the interrelationship among six characteristics. In Figure 3, we illustrate the result in accordance to the MST approach and Pajek software was used to illustrate the network among those characteristics.

Figure 3 represents the network for robust and non robust approach, respectively. We can see clearly that we have two difference interrelationship (or network). Consequently, according to the centrality measures, we will achieve difference results of centrality. However, the non robust could be meaningless since their network have the outlier (before correction). Therefore, we only consider the robust approach for further analysis (after correction).

3.2 Centrality Measure: The Influential Characteristics

Based on degree centrality, the higher the number of links represents the greater of the role of the respective characteristic in influencing the others. As a results in Table 3, it could be concluded that Malaysian universities and colleges should pay more attention on Repurchase 2 - I would choose my university for another degree since it has the higher number of links (3), followed by Service quality and Repurchase 1 (2 links), whereas the remaining characteristics; Word mouth to mouth and Good experience have only one (1) link.
Table 3. Centrality measure

<table>
<thead>
<tr>
<th>Node</th>
<th>Characteristic</th>
<th>Degree</th>
<th>Average of weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Good experience</td>
<td>1</td>
<td>0.7808</td>
</tr>
<tr>
<td>2</td>
<td>Service quality</td>
<td>2</td>
<td>0.8004</td>
</tr>
<tr>
<td>3</td>
<td>Repurchase 1</td>
<td>2</td>
<td>0.5952</td>
</tr>
<tr>
<td>4</td>
<td>Repurchase 2</td>
<td>3</td>
<td>0.7166</td>
</tr>
<tr>
<td>5</td>
<td>Word mouth to mouth</td>
<td>1</td>
<td>0.6964</td>
</tr>
<tr>
<td>6</td>
<td>Pay premium price</td>
<td>1</td>
<td>0.8359</td>
</tr>
</tbody>
</table>

However, our proposed method shows that *Pay premium price* - *I would consider making a financial gift to my university* is the key point to the university management for improving the students’ satisfaction. In this measure, the higher the ‘average of weights’ indicates the more important is their influence as compared to the other characteristics. Even though, in degree centrality, *Pay premium price* has only one (1) link, it appears to be the most important characteristic since we take into consideration its “contribution” rather than its “popularity”.

4. Conclusion

In this research, we are able to identify the influential characteristics using network analysis. The robust approach was successfully deployed to develop a correlation matrix to investigate the interrelationship among those characteristics, especially when involving a high dimension dataset.

Six characteristics and their correlation structure are considered as a network system. To simplify that network, we used minimum spanning tree which provide an optimal sub-network in the form of a spanning tree. This tree is then used to construct the optimal network analysis of those characteristics.

We also introduced “average of weights” as a new centrality measure. Its advantage is illustrated by using six characteristics of student’s satisfaction, together with the other established centrality measures, degree centrality. It could help to enrich the interpretation of correlation network. According to the centrality measures, the educational institutions have to focus on *Pay premium price* since it reflects the highest degree of satisfaction for the international students. This finding is consistent with studies by Anderson et al. (1997) and Yeo (2008).

This study renders two potential research directions. First, this study shows that a centrality measure can help us to identify the most influential characteristics. Based on degree centrality and average of weights centrality, the characteristics (nodes) that have only 1 link is actually can be suspected as an outlier. Hence, future research can be devoted into identifying outlier by using the notion of MST. Second, in this paper we use a Kruskal’s algorithm who gives us one possible network only. However, the use of MST as an information filter might produce misleading information if there is more than one MST in the network. More studies can be carried out to investigate the limitation of this topic.

References


96


